



Beyond the Implicit/Explicit Dichotomy: The Pragmatics of Plausible Deniability

Francesca Bonalumi¹ · Johannes B. Mahr² · Pauline Marie³ ·
Nausicaa Pouscoulous³

Accepted: 28 August 2023
© The Author(s) 2023

Abstract

In everyday conversation, messages are often communicated indirectly, implicitly. Why do we seem to communicate so inefficiently? How speakers choose to express a message (modulating confidence, using less explicit formulations) has been proposed to impact how committed they will appear to be to its content. This commitment can be assessed in terms of accountability – is the speaker held accountable for what they communicated? – and deniability – can the speaker plausibly deny they intended to communicate it? We investigated two factors that may influence commitment to implicitly conveyed messages. In a preregistered online study, we tested the hypothesis that the degree of meaning strength (strongly or weakly communicated) and the level of meaning used by the speaker (an enrichment or a conversational implicature) modulate accountability and plausible deniability. Our results show that both meaning strength and level of meaning influence speaker accountability and plausible deniability. Participants perceived enrichments to be harder to deny than conversational implicatures, and strongly implied content as more difficult to deny than weakly implied content. Furthermore, participants held the speaker more accountable to content conveyed via an enrichment than to content conveyed via an implicature. These results corroborate previously found differences between levels of meaning (enrichment vs. implicature). They also highlight the largely understudied role of meaning strength as a cue to speaker commitment in communication.

1 Introduction

An important challenge in understanding human communication is the question of what processes drive message construction. Why do speakers construct utterances in the way they do? How do they generate utterances to achieve their intended effects?

Francesca Bonalumi and Johannes B. Mahr contributed equally to this work.

Extended author information available on the last page of the article

Recently, speaker commitment emerged as fundamental in answering these questions (Geurts 2019): utterances commit the speaker to the truth of a proposition or to a future course of action (Hamblin 1971; Beyssade and Marandin 2009), but such commitments are only effective if listeners are able to track these commitments and hold speakers accountable to them (e.g., Vullioud et al. 2017; see also Mahr and Csibra 2018, 2020, 2021).

How speakers choose to express their message should impact how committed they are perceived to be – how much others will hold them *accountable* to what they have expressed and the degree of *plausible deniability* of their message. Speakers are held accountable for the message they convey and will incur reputational costs if it is found to be unreliable (Boulat and Maillat 2017; Brabanter and Dendale 2008). It is therefore sometimes in their interest if this message is conveyed in such a way that it is plausibly deniable. While the notion of plausible deniability is first a psychological one, which relies on intuitive judgments, several theoretical analyses have now been put forward to account for it (Camp 2018, 2022; Dinges and Zakkou 2023; Mazarrella 2021; Pinker 2007; Pinker et al. 2008). Plausible deniability allows speakers to refute having intended a certain message (typically an implicit one), for instance when confronted by the recipient (Brown and Levinson 1987; Lee and Pinker 2010; Pinker 2007; Pinker et al. 2008). By taking a strategic approach to utterance construction, speakers can manipulate the extent to which the audience can hold them accountable for the meaning they have conveyed (Pinker et al. 2008; Soltys et al. 2014). Illustrating this, Lee and Pinker (2010) found that speakers favoured implicit constructions when they were asked how they would attempt to bribe a policeman. The choice to forego an explicit offer in such a scenario is strategic, as speakers can deny their implicit offer and avoid unpleasant and/or awkward social repercussions. Speakers can thus mitigate the risks of a negative outcome, since ‘cooperative’ recipients can accept implicit offers, while ‘antagonistic’ recipients would not have enough evidence to confront them. Therefore, how committed speakers are perceived to be is importantly related to plausible deniability. On the one hand, deniability might influence how speakers construct their utterances in strategic situations. On the other hand, plausible deniability might be a function of how strongly a speaker is committed.

However, while multiple factors have been found to influence perceived speaker accountability, much less work has focused on the factors that determine plausible deniability beyond how explicitly or implicitly a message is conveyed. One factor research has focused on is speaker’s confidence in expressing a message, which can be marked either linguistically (e.g., ‘I guess’, ‘I’m sure’) or non-verbally through posture, gestures, tone of voice or facial expressions. Speaker confidence has been found to impact both the message’s credibility – by increasing its chances of being accepted by the interlocutor – and the speaker’s accountability – namely the social repercussions if the message is found to be unreliable (Vullioud et al. 2017; Mazarrella et al. 2018). Similarly, claims about the source of one’s information (“I saw it” vs. “Somebody told me”) have been shown to have an impact on the message’s credibility, as well as the speaker’s accountability (Mahr and Csibra 2021).

Here we extend this work, by focusing explicitly on two additional pragmatic factors that might not only influence speaker accountability but also plausible deniability.

ity: the level of meaning and its strength. First, speakers can convey messages with different degrees of explicitness, that is, using different levels of meaning: speakers communicate not only the linguistically encoded meaning of their utterances, they can also pragmatically *enrich* the content of these utterances or *implicate* propositions in addition to what they explicitly say (Grice 1989). The meaning of the conjunction ‘and’, for instance, can be pragmatically enriched to ‘and then’. An utterance of “they got married and had a child” will then prompt the hearer to believe the two events happened in a certain order. The meaning of an utterance can depart even more from its linguistic form and its explicit content through an implicature, for instance when a child’s request for a snack is met with an implicit denial in the form of “dinner will be ready in 10 minutes”.

Second, speakers can convey messages with different degrees of manifestness – e.g., a pragmatically inferred part of meaning (be it an enrichment or an implicature) may be more or less strongly communicated. Meaning strength is conceived here as the accessibility of what is communicated; it depends both on how manifest the speaker made their intention to communicate a specific content and how important (or inconsequential) the recovery of this content is for the interpretation of the utterance (Wilson and Sperber 2004). Meaning strength depends on the context of the utterance. For instance, “they got married and had a child” is likely to strongly communicate that the marriage preceded the child’s birth in the context of a discussion about the couple’s adherence to certain social norms. On the other hand, if the same sentence appears in an exchange about a drop in the couple’s social engagements, the temporal reading of *and* might still be accessible, but is only weakly communicated. Importantly, both the level of meaning and the strength with which a content is expressed may have an impact on how committed a speaker will be perceived to be.

Given that a speaker risks suffering social repercussions when implying something false, would these repercussions be more severe when this content is communicated via a pragmatic enrichment rather than via an implicature? Or would they be worse off if the false content is strongly rather than weakly implicated? Alternatively, would the differences in speaker accountability emerge only when these different types of contents – weakly or strongly implicated or enriched – are denied? Here, we aimed to investigate the influence of the level of meaning and meaning strength on the speaker’s accountability, and whether they bring about similar effects on their message’s plausible deniability.

1.1 How might the Level of Meaning Impact Accountability and Plausible Deniability?

It has recently been proposed that commitment is stronger when meaning is fully linguistically encoded (explicitly communicated) than when it is merely implicated (Morency et al. 2008; Reboul 2017). Indeed, conversational implicatures have been found to foster less accountability than explicit contents, while explicit contents and presuppositions lead to similar levels of accountability (Hall and Mazzarella 2023; Mazzarella et al. 2018). Plausible deniability should be similarly influenced by the degree of explicitness of the conveyed meaning – i.e., by its level of meaning (Brown and Levinson 1987).

Pragmatic inferences are by definition cancellable – i.e., an utterance that implies a proposition p can be followed by the phrase ‘but not p ’ without logical contradiction (Grice 1989; Levinson 2000). However, a proposition may be cancellable without being deniable – or at least, not plausibly so (Pinker et al. 2008): you may be able to logically cancel the non-literal content of your utterance without being in a position to plausibly deny intending it in the first place. If, when asked whether you know about a new policy introduced at a meeting, you answer “I wasn’t there” the implication that you did not know is clearly cancellable – there is no logical contradiction between not being present and knowing about something. Yet, it will be hard to deny that you intended to convey your ignorance of the new policy. Pragmatic phenomena can therefore be logically cancelled but whether, and how, they can be plausibly *denied* is a more complex issue that will depend largely on the properties of the context of utterance (Mazzarella 2021).

Of course, denying that some fully explicit content was intended is difficult – short of lying or claiming a mistake. The phenomenon of plausible deniability primarily applies to contents conveyed via pragmatic inferences, which seem much easier to revoke. Yet, these do not form a homogeneous category: the outcome of a pragmatic inference might be more or less explicit and thus, we believe, more or less plausibly deniable.

Since Grice (1989), most pragmatists consider there to be different types of implicit (i.e., cancellable) contents. Pragmatic enrichments,¹ such as the temporal reading of *and*, go beyond what is linguistically encoded in an utterance, but they are linked to linguistic terms or structures, in the presence of which they will often be derived (e.g., conjunction ‘and’ enriched to ‘and then’; disjunction ‘or’ enriched to ‘not both’; quantifier, ‘some’ or ‘most’ enriched to ‘not all’). In contrast, implicatures², such as the implicit denial for a snack seen above, are entirely context-dependent additional propositions and do not depend on any specific linguistic feature. Their derivation requires taking into account the proposition explicitly expressed, as well as the context of utterance and the speaker’s intention (see e.g., Grice 1989; Levinson, 1983; Sperber and Wilson 1986/95).³ Pragmatic enrichments are considered more closely related than implicatures to the linguistic form and the explicit content of an utterance: some theorists maintain that they are derived automatically, unless blocked by the context (Horn 1989; Levinson 2000), while others argue that, despite being

¹ We will refer to pragmatic inferences linked to scalar terms, quantifiers, modals, numerals, or to logical terms (such as conjunction) as pragmatic ‘enrichments’ throughout the paper – following authors such as Récanati (2004), Carston (2002) and Sperber and Wilson (1986/95). These phenomena are considered to be ‘Generalised Conversational Implicatures’ (GCIs) by Gricean (Grice 1989) and neo-Gricean theorists (e.g., Horn 1989; and Levinson 2000). Although the difference in terminology corresponds to important differences in how different pragmatic theorists view these phenomena, these have no direct bearing on our study.

² Similarly, we will refer to the pragmatic inferences corresponding to ‘Particularised Conversation Implicatures’ (PCIs) in Grice’s (1989) terminology simply as ‘implicatures’. Note that for most of the relevant experimental literature it would be safe to assume that ‘enrichments’ and ‘implicatures’ correspond to the same phenomena as, respectively, GCIs and PCIs.

³ On the thorny issue of how to establish whether a pragmatic phenomenon is an enrichment or an implicature, see Carston 2002. Whether these two sets of pragmatic phenomena are derived through the same cognitive processes has also given rise to much debate (e.g., Récanati 2004).

pragmatically inferred, they are part of the explicit content of the utterance ('what is said', in Gricean terms) and contribute to its truth conditions (Carston 2002; Levinson 2000; Récanati 1993, 2001, 2004; Sperber and Wilson 1986/95). Theoretical debates notwithstanding, there is a general sense that pragmatic phenomena might be more or less explicit, with implicatures firmly lodged in the implicit camp and pragmatic enrichments verging toward, or achieving, explicitness (see, for example, Levinson 2000).

A large amount of research focuses on whether participants can distinguish enrichments from explicit content ('what is said' in Gricean terms) to determine whether they contribute to the truth-conditional meaning of the utterance. Enrichments were originally found to be judged as part of 'what is said' (Gibbs and Moise 1997), but the picture subsequently became more complex (Bezuidenhout and Cutting 2002). It seems that participants' intuitions differ depending on the type of pragmatic inference (Doran et al. 2009) and the task used – for instance, the types of enrichments investigated by Doran and colleagues (2012) were neither consistently included nor excluded from 'what is said'.

Relatedly, several recent studies investigated whether interlocutors consider false information conveyed via pragmatic inferences to be an instance of lying. This research on the comprehension of deceitful implicatures and enrichments yields mixed findings (for a study on the production of misleading enrichment and implicature, see Franke et al. 2020).

Weissman and Terkourafi (2019) show that participants consistently judged false implicatures to be non-lies, while some types of enrichments (e.g., upper-bounded interpretation of numerals) were more easily considered to be lies when the inferred meaning was false. Hall and Mazzarella (2023) also found that speakers suffered more reputational cost when they communicate false information via enrichments than when they communicated the same information via implicature. Yet, other studies suggest that even false implicatures may be considered eligible to be lies, and sometimes as much so as false enrichments (Antomo et al. 2018; Viebahn et al. 2018; Willemsen & Wiegmann, 2017). Furthermore, Wiegmann (2022) showed that many among the implicatures used by Weissman and Terkourafi (2019) were still considered to be cases of lying (and, thus, the speaker presumably conceived to be committed to the implicated content) provided that the intention to communicate the implicated (and deceptive) content was made salient.

Notwithstanding the fact that there are likely cultural differences in people's intuition of what counts as a lie (e.g., Danziger 2010; Hardin 2010; Hruschka 2020), Reins and Wiegmann (2021) used a variety of particularly relevant measures to investigate the folk notion of lying. Following four scenarios involving a false implicature or enrichment, participants were asked whether they considered them to be lies, among other questions; these were compared to an additional set of measures including (among others) explicit questions about commitment (did the speaker commit themselves to the false enrichment/implicature) and deniability (could the speaker convincingly deny it). Lie responses correlated with those assessing commitment and deniability. Reins and Wiegmann found that false implicatures were mostly judged to be lies. However, attributions of lying, as well as commitment and deniability, were lower for implicatures than for enrichments. In a similar vein, Bonalumi

and colleagues (2020) found that people considered unfulfilled promises conveyed via enrichment to have been broken, but not those conveyed via implicature.

Overall, although enrichments are pragmatic inferences – and thus cancellable – they often seem to be perceived as contributing to the utterance’s truth-conditions. As a result, they generally appear more difficult to deny than implicatures, which – it is generally agreed – are part of ‘what is implicated’. Additionally, as mentioned earlier, implicatures lead to less accountability than the explicit content of an utterance (Hall and Mazzearella 2023; Mazzearella et al. 2018). The two levels of meaning – enrichments and implicatures – should, therefore, modulate both accountability and plausible deniability differently (as suggested by Bonalumi et al. 2020; Reins and Wiegmann 2021; Wiegmann 2022).

1.2 How might Meaning Strength Impact Accountability and Plausible Deniability?

Another factor that can be linked to both plausible deniability and accountability is the degree of manifestness of what is communicated, i.e., its strength⁴. A pragmatically inferred part of meaning (be it an enrichment or an implicature) can be more or less strongly communicated; this will depend on how manifest the speaker made their intention to communicate it (Wilson and Sperber 2002), as well as how essential the pragmatic content is to understand the overall communicative act. First, a strongly pragmatically inferred content (enrichment or implicature) is a proposition the speaker intends to communicate, and they will therefore make this intention clear to their interlocutor, whereas their intention to communicate weaker contents is hazier and, as a result, less manifest to their interlocutor. Second, a strong implicature or enrichment is generally crucial to make the speaker’s utterance relevant in context, while the recovery of a weak one might be optional. Both aspects point towards strongly pragmatically inferred parts of meaning as being more accessible for the hearer than weaker ones.

Imagine two young parents arriving at home after braving a downpour and commenting on how they never want to leave the comfort of their living-room again. If one of them exclaims: “We’re out of milk!” the implicature that someone must go out to get milk for the child is strongly communicated. Indeed, it is difficult to see how the utterance would be relevant in this context if the implicature was not intended. On the other hand, other (weaker) implicatures might have been intended, or not – e.g., “our child drinks more milk than she used to” or “you forgot to put milk on the grocery list again”⁵.

As Mazzearella (2021) notes, the strength of an implicit meaning will affect who endorses responsibility for the pragmatic inference. In the case of a strong implicature

⁴ The notion of implicature strength was introduced by relevance theorists (Sperber and Wilson 1986/1995; 2004; Wilson and Sperber 2002), yet we believe it to be a very useful tool in the analysis of accountability independently from the specific theoretical apparatus of Relevance theory.

⁵ Similarly, in a different context, for instance uttered by parents of older children, “We’re out of milk!” might only weakly communicate the implicature that someone must go out and buy milk – as it may not convey this implicature at all in other contexts, such as if uttered on a camping site miles away from any shop.

or enrichment, the speaker bears more responsibility since they make their intention to communicate it clearly manifest. Inversely, in the case of a weak implicature or enrichment the responsibility of deriving it lies mostly with the hearer (Sperber and Wilson 2006). This, in turn, should involve consequences for how accountable the speaker will be perceived to be: the stronger the implicature or enrichment, the more committed to it the speaker should appear (see also Boulat and Maillat 2023). Plausible deniability should be equally affected by meaning strength. Since the derivation of a strong implicature or enrichment is paramount to understanding the utterance, there is little room left in these cases for an alternative interpretation. Any attempt of denial would, thus, be less plausible, since there is only a narrow range of possibilities to re-construct the context – and thus providing an alternative, non-committal, interpretation of the utterance (Mazzarella 2021). On the other hand, weak implicatures and enrichments offer the speaker exactly this range of possibilities, suggesting that meaning strength is inversely connected to plausible deniability.

The hypotheses that stronger implicatures and enrichments should be more accessible than weaker ones, but also more committal, are mostly borne out by the handful of studies investigating meaning strength. Nicolle and Clark (1999) first found that strong implicatures prompted participants to select the implied meaning conveyed by the utterance as the best reflection of ‘what [it] said’. In contrast, with weaker implicatures, participants were more likely to select the minimal proposition of the utterance as representative of the explicit content (‘what is said’). Consistent with the hypothesis that meaning strength (modulated by relevance) influences accountability and plausible deniability, Bonalumi and colleagues (2020) found that the same explicit broken promise produced different social repercussions for the speaker depending on whether the recipient was known to rely on the promise made by the speaker. Finally, Sternau et al. (2015) investigated the deniability of enrichments, as well as weak and strong implicatures using an explicit question about deniability (akin to Reis & Wiegmann, 2021). Although their findings rely on participants’ *a priori* intuitions, rather than actual attempts of denial, they indicated that enrichments are perceived as less deniable than implicatures, and strong implicatures less so than weak ones.

Taken together, these results suggest that strong implicatures are more easily included into the explicit content, have higher impact on accountability and might be harder to deny compared to weaker ones. Note that since meaning strength is a feature of enrichments and implicatures alike (Clark 2013), it should modulate commitment and plausible deniability for both phenomena (Mazzarella 2021; and Sternau et al. 2017 both also make this prediction for plausible deniability).

1.3 The Present Study

A message can be conveyed both with different degrees of manifestness, i.e., more or less strongly, and with different degrees of explicitness, i.e., graded levels of meaning (see Fig. 1). As reviewed above, both factors might affect the extent to which a speaker is committed to the content of her utterance, and consequently how easily that commitment is deniable. Here, to further understand the attribution of commitment (and how speakers strategically attempt to avoid it), we aimed to directly test

the effects of meaning strength and level of meaning on deniability. Specifically, we sought to test the following hypotheses:

1. Strongly implicated contents should lead to higher accountability and be more difficult to deny than weakly implicated contents.
2. Enrichments should lead to higher accountability, and be more difficult to deny, than implicatures.

We tested these hypotheses in an online experiment. The experiment itself – the Commitment Experiment – was preceded by a Norming Study to ensure participants interpreted the materials with the pragmatic inferences and drew comparable intended meanings in each experimental condition. Participants were presented with written scenarios in which speakers used different levels of meaning (*Enrichments* or *Implicatures*) with different strengths (*Weak* or *Strong*) to make a commitment. Meaning strength depended on how manifest the implied content was in the scenario. Implicatures were always determined by the interaction between the utterance and the specific context, while enrichments were of four types: temporal conjunction enrichment (‘and’ meaning ‘and then’), the enrichment of the conditional (‘if’) to a biconditional (‘if and only if’), as well as the scalar inferences linked to the quantifier

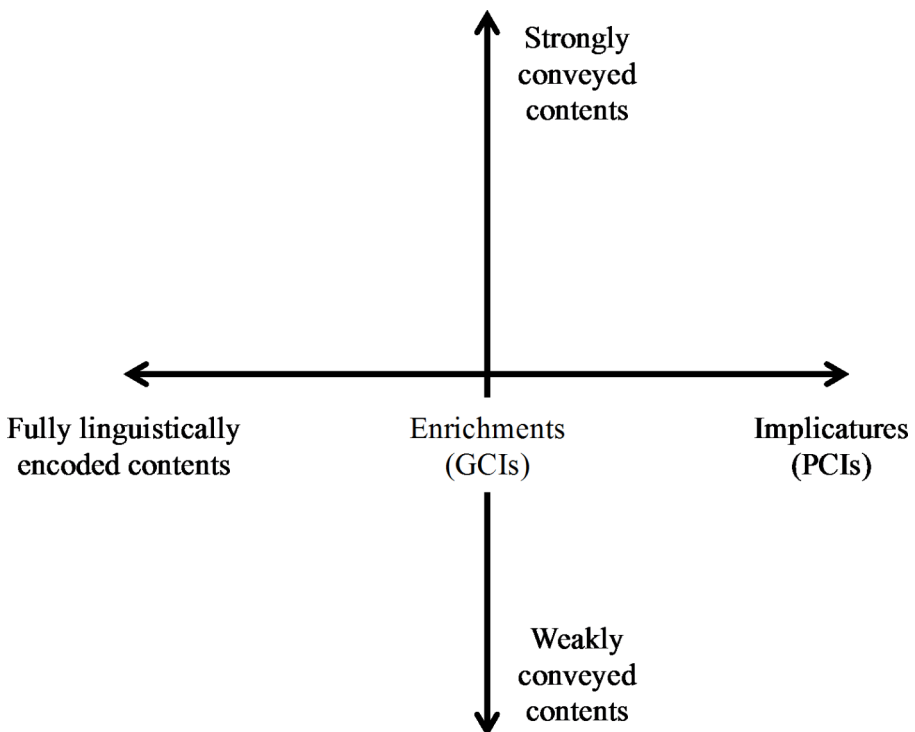


Fig. 1 The degree of explicitness (x-axis) interacts with the degree of manifestness (y-axis): fully linguistically encoded contents, as well as enrichments and implicatures, can be more or less strongly/weakly communicated

‘some’ (‘not all’) and to disjunction (‘or’ meaning ‘not both’; for discussion of these phenomena see for instance, Levinson 2000; Carston 2002; Noveck 2004). We chose four different types of enrichments to make sure our findings could generalise to enrichments as a category and were not limited to a single phenomenon. We took care to pick phenomena generally agreed to be enrichments across theoretical frameworks – i.e., all of the phenomena we chose are considered as enrichments or as generalised conversational implicatures (GCIs) by authors such as relevance theorists (Sperber and Wilson 1986/95; Carston 2002; Noveck and Sperber 2007), Récanati (2004), and by neo-Griceans (such as Levinson 2000). These authors agree that these pragmatic phenomena differ from (particularised) implicatures and may (unlike implicatures) contribute to the truth-conditions we intuitively assign to the utterances in which they appear.

In the Norming Study, we measured whether participants inferred the expected enriched/implicated meaning by asking them whether the speaker intended it, as well as their level of confidence about their answer. The Commitment Experiment itself was conducted to test our main hypotheses that more strongly implicated contents, on the one hand, and enrichments, on the other, lead to higher accountability and are harder to deny than weaker implicated contents and implicatures, respectively. Participants were presented with scenarios (selected from those tested in the norming study) in which speakers later broke their implied commitment. Following previous studies investigating speaker commitment (Mazzarella et al. 2018; Bonalumi et al. 2020; Yuan and Lyu 2022), we used two indirect measures to assess whether the speaker was held accountable for breaking their commitment: asking participants whether the speaker was blameworthy (which relates to partner control attitudes, i.e., judgements aimed to manipulate partners’ behaviours so to disincentivise these behaviours in the future) and trustworthy (which relates to partner choice attitudes, i.e., judgements indicating one’s willingness to avoid the interaction with that partner in the future). Previous results suggest that trustworthiness and blameworthiness judgments tend to mirror each other (although the former seem to be more influenced by pragmatic factors than the latter, as shown by Mazzarella et al. 2018). We, therefore, predicted that these two measures would provide similar results: a speaker judged more blameworthy would be deemed less trustworthy, and vice versa. Additionally, the influence of the level of meaning and of meaning strength on denial was also assessed. There is always a risk that participants’ *a priori* intuitions on something like deniability (as measured by Reins and Wiegmann 2021 and Sternau et al. 2015) would be coloured by pre-assumptions or other kind of judgements about deniability—hypothetical considerations may generate additional inferences that colour participant answers and may not accurately reflect acceptance of an actual denial. We therefore favoured a more direct approach to establish whether an implied content is perceived as plausibly deniable or not: we compared participants’ judgments following the broken commitment with, and without, denial. A decrease of accountability following denial would indicate the plausible deniability of the implied content; if speakers making an implied commitment are held less accountable when it is denied compared to when it is not denied, it is fair to conclude that the implied commitment was (to some extent) deniable. We predicted that the impact of meaning strength

and of the level of meaning on plausible deniability would mirror their impact on accountability.

2 Norming Study

The Norming Study focused on the derivation of the appropriate pragmatic meanings, conveyed either by enrichments or by implicatures in different contexts. The results of this study were then used to make informed decisions about which scenarios to use in the main experiment. Participants were not presented with any breach of the implied commitment, since this might have confounded participants' responses to the implicature question.

2.1 Methods

2.1.1 Participants

We recruited 150 participants (105 females, $M_{age} = 33.23$) through Prolific Academic (Palan and Schitter 2018). Eligibility criteria for participation in the study were age (20 to 70) and first language (English). We excluded participants if their completion time diverged by more than three standard deviations from the mean ($M = 379.86$ s, $SD = 375.6$ s; $N = 2$), or if they provided wrong answers to the comprehension question ($N = 29$). Participants provided informed consent before taking part in the experiment and were paid £0.84 for their time.

2.1.2 Materials

We created 21 scenarios, each with the following structure:

- Context describing the situation in which the utterance occurs – manipulated so to create *Weak* and *Strong* conditions;
- Dialogue containing the speaker's implied commitment to accomplish a specific task. The commitment was conveyed either via *Enrichment* or *Implicature* (see Table 1).

All the scenarios are available at <https://osf.io/2fu93/files/osfstorage>.

We modulated meaning strength through changes in the scenario context resulting in an increase or decrease of the relevance of the speaker's utterance. This change in relevance should affect the accessibility of the implied commitment. For instance, in the scenario mentioned in Table 1, the change in context would concern the number of customers waiting to be served. As Elliot utters, "I'll finish this cocktail and change the keg," in a weak context where few customers are waiting, the order in which he does so should be less relevant than in the context in which there are numerous customers to serve. Although the implicature would be drawn in both contexts, in the latter, stronger context, Sophie would understand Elliot's utterance as more pointedly ordered, thus resulting in a more accessible and stronger implicature.

Table 1 Examples of scenarios in the four conditions

<i>Enrichment condition</i>	
<i>Weak condition</i>	<i>Strong condition</i>
Sophie and Elliot are colleagues and both work in a bar as bartenders. There is no more craft beer on tap and Sophie and Elliot have to change the keg. It is a Tuesday night and there are very few customers. <u>Sophie</u> : There's no more craft beer on tap. <u>Elliot</u> : I'll finish making this cocktail and change the keg.	Sophie and Elliot are colleagues and both work in a bar as bartenders. There is no more craft beer on tap and Sophie and Elliot have to change the keg. It is a Saturday night and there are a lot of impatient customers.
<i>Implicature condition</i>	
<i>Weak condition</i>	<i>Strong condition</i>
Scott and Samy meet every week to watch the new episode of their favourite TV show together. This week, Samy can't make it. When this happens, the one who can't make it usually catches up on the missed episode before the next one the following week. <u>Samy</u> : I won't be able to come over to watch the show with you this week. I have an urgent deadline. <u>Scott</u> : It's ok, there's a new movie I've been meaning to see.	Scott and Samy meet every week to watch the new episode of their favourite TV show together. This week, Samy can't make it. It is the first time it happens and Samy is upset he can't make it.

The level of meaning was examined using 8 scenarios conveying commitment via an implicature and 13 scenarios conveying commitment via an enrichment. Four types of enrichments were used: two types of scalar implicatures – one linked to the quantifier ‘some’ (‘some’, but ‘not all’), the other to the connective ‘or’ (but ‘not both’) – as well as the temporal enrichment of the connective ‘and’ (to ‘and then’) and conditional perfection, where ‘only if’ is derived from ‘if’ (see Table 2 for examples).

2.1.3 Procedure and Design

The Norming Study used a mixed factorial 2 (‘strength’ as between-subjects factor: *Strong* vs. *Weak*) x 2 (‘level of meaning’ as within-subject factor: *Enrichment* vs. *Implicature*) design. Participants were randomly presented with four enrichments and two implicatures and saw all of them either in a strong or in a weak version. After reading each scenario, participants answered a comprehension question about the scenario content. Participants were then reminded of the speaker’s utterance containing the implicit commitment, and were asked to answer an Implicature question and a Confidence Rating question, as illustrated below:

- **Implicature Question**

Do you understand Elliot [*speaker*] to have meant that he will finish making the drink first, and then change the keg [*implied commitment*]?

(Yes/No/I don't know)

- **Confidence Rating**

How confident are you in your answer?

Rate your confidence from 0 to 5, with 0 being ‘not at all’ and 5 being ‘completely’. (6-point Likert scale)

Table 2 Examples of different types of enrichment used in the norming study and in the commitment experiment

Type of enrichment	Example	Intended meaning
<i>Scalar implicature quantifier 'some'</i>	I'll take some tissue packets with me to the office today.	I'll take some tissue packets with me to the office today, but not all the tissue packets.
<i>Scalar implicature connective 'or'</i>	Yes, I'll need the projector or the eraser board.	Yes, I'll need either the projector or the eraser board, but not both .
<i>Conjunction enrichment of 'and'</i>	I'll finish making this cocktail and change the keg.	I'll first finish making this cocktail and then change the keg.
<i>Conditional perfection with 'if'</i>	If you don't get a dog then I'll get a cat	I'll get a cat, if and only if you don't get a dog.

2.2 Results and Discussion

We took the answers 'No', and 'I don't know' to the implicature question as indicating that the implied meaning had not been inferred by participants, while 'Yes' answers were interpreted to indicate that the implicature had been drawn. To get a measure of how reliably participants inferred the implied meaning, we converted participants' responses to the implicature and confidence questions into a unified 'inference score' by multiplying participants' confidence responses by -1 when participants failed to draw the correct inference and by +1 when they succeeded (for a similar approach, see Starmans and Friedman 2012).

Based on this measure, we selected the four implicature scenarios with the highest scores in both strong and weak contexts. The same process was applied for the enrichments (on 'and', 'or', 'some' and 'if'), we chose the scenario with the highest inference score. However, in the case of 'or' enrichments, we found the scenario with the highest inference score to be flawed: the inference giving rise to the exclusive meaning was a product of world knowledge rather than of the utterance in context. Therefore, the enrichment with the second-highest inference score in this category was selected instead. Across all selected scenarios, 95.79% of participants drew the correct pragmatic inference and the average inferences score was 4.21 out of 5.

3 Commitment Experiment

The Commitment Experiment was designed to test our main hypothesis that meaning strength and level of meaning modulate both the extent to which a speaker is held accountable for their implied commitment and the extent to which a speaker can plausibly deny such a commitment. We predicted that contents conveyed via enrichment would lead to higher accountability ratings and would be less easily deniable than contents conveyed via implicature. We also predicted that strongly conveyed contents would lead to higher accountability ratings and would be less easily deniable than weakly conveyed contents.

We measured perceived accountability through two test questions: one asking how blameworthy the speaker is (blame question), the other asking how much participants would trust the speaker in the future (mistrust question). The blame question

measured commitment attribution via blameworthiness; the mistrust question measured whether participants would draw any inference about the speaker's dispositions (resulting from the breach of the commitment) which may be relevant to future interactions. While accountability was measured using direct questions, to assess the effect of plausible deniability we introduced an additional between-subjects factor: the denial of the implied commitment. If the accountability ratings of an implied commitment are lower when it is denied than when it is not, then the implied commitment was (at least to some extent) deniable.

3.1 Methods

The study was pre-registered on OSF.io, with the sample size, planned analyses and participant exclusion criteria specified. The pre-registration document is available at <https://osf.io/nkv63/>. Some of the analyses presented in the paper deviate from the pre-registered ones in view of comments received during reviewing processes and conference presentations.

3.1.1 Participants

A power analysis conducted with G*Power 3.1 (Faul et al. 2009) revealed that, assuming a small to medium effect size ($F=0.175$), a sample size of 400 participants would convey a statistical power of 0.89. Consequently, we recruited 400 participants (297 females, $M_{age} = 35.39$) through Prolific, with age (between 20 and 70) and first language (English) as eligibility criteria. Participants provided informed consent and were compensated with £0.09. Participants were exposed to only one story, and each experimental session lasted few minutes (an average of 112s to place a response). Data was discarded for participants who took too much or too little time to complete the study (three standard deviations from the mean; $N=7$) or failed the comprehension question ($N=6$), totalling 13 excluded participants. The final analysis thus included 387 participants.

3.1.2 Materials

The commitment experiment used the eight scenarios selected based on the results of the Norming Experiment. In four scenarios the commitment was conveyed through an implicature and in four others the commitment was conveyed through an enrichment, one in each category: quantifier 'some', connectives 'or' and 'and', conditional 'if' (see, Table 2).

In contrast to the Norming Experiment, the commitment experiment included an additional between-subjects factor: 'denial'. In the *Denial Present* condition, the breach of the commitment was followed by the speaker's denial of the implied content, when confronted by the recipient. The denial always appeared in the following form: "I didn't say I would [*implied commitment*]. I said I would [*as explicitly stated*]".

The scenarios thus had the following structure (see Table 3 for an example):

- Context describing the situation in which the utterance occurs – manipulated so to create *Weak* and *Strong* conditions.
- Dialogue containing the speaker’s implied commitment to accomplish a specific task – conveyed either via *Enrichment* or via *Implicature*.
- Breach of the implied commitment.
- Dialogue containing the speaker’s denial of the implied commitment – only the participants in the *Denial Present* condition were presented with this dialogue.

In contrast to the Norming Experiment, there was no implicature question in the commitment experiment, since the presence of denial in half of the conditions would have confounded the responses to the implicature question.

3.1.3 Design and Procedure

The commitment experiment had three between-subjects factors: ‘strength’ (*Strong* vs. *Weak*), ‘level of meaning’ (*Implicature* vs. *Enrichment*), and ‘denial’ (*Denial Present* vs. *Denial Absent*), resulting in eight different experimental conditions. Participants were randomly assigned to one of the eight conditions, and each participant was only exposed to one scenario in one condition. Participants were told to read the scenario, and were then presented with a multiple-choice comprehension question. Participants were subsequently reminded of the speaker’s utterance containing the implied commitment and were presented with a blame question and a mistrust question on 6-point Likert scales (see Table 3 for an example).

We expected the blame question and the mistrust question to show similar patterns of results. For the blame question, we predicted that participants would be more likely to blame the speaker in the *Strong* condition than in the *Weak* condition, and that they would be more likely to blame the speaker in the *Enrichment* condition than in the *Implicature* condition. Similarly, for the mistrust question, we predicted that participants would be more likely to mistrust the speaker in the *Strong* condition than in the *Weak* condition, and that they would be more likely to mistrust the speaker in the *Enrichment* condition than in the *Implicature* condition.

Moreover, we predicted interactions between meaning strength and the presence of denial, as well as between level of meaning and the presence of denial. First, we expected weakly (but not strongly) conveyed contents to be more plausibly deniable. Therefore, we predict that in the *Weak* condition (but not in the *Strong* condition) participants would be less likely to blame and to mistrust the speaker in the *Denial Present* condition compared to the *Denial Absent* condition. Second, we expected implicatures (but less so enrichments) to be more plausibly deniable. Therefore, we predicted that in the *Implicature* conditions (but not in the *Enrichment* condition), participants would be less likely to blame and mistrust the speaker in the *Denial Present* condition compared to the *Denial Absent* condition.

Table 3 Scenario structure and measures used in the commitment experiment with example (of the *Strong+Enrichment+Denial Present* conditions)

	<i>Example</i>
<i>Context</i> (either <i>Strong</i> or <i>Weak</i>)	Sophie and Elliot are colleagues and both work in a bar as bartenders. There is no more craft beer on tap and Sophie and Elliot have to change the keg. It is a Saturday night and there are a lot of impatient customers.
<i>Implied commitment</i> (either <i>Enrichment</i> or <i>Implicature</i>)	<u>Sophie</u> : There's no more craft beer on tap. <u>Elliot</u> : I'll finish making this cocktail and change the keg.
<i>Breach of implied commitment</i>	Elliot leaves the cocktail he was making and goes to change the keg. The customer whose cocktail it was complains to Sophie. Sophie is unhappy about this.
<i>Comprehension Question</i>	<i>Where are Sophie and Elliot working?</i> <ul style="list-style-type: none"> • <i>A café</i> • <i>A pub</i> • <i>A restaurant</i> • <i>An ice-cream parlour</i>
<i>Denial of implied commitment</i> (either <i>Present</i> or <i>Absent</i>)	<u>Sophie</u> : You said you would finish making the cocktail before changing the keg! <u>Elliot</u> : I didn't say that I would do that first. I said I'd finish making the cocktail and change the keg.
<i>Utterance reminder</i>	Remember Elliot said: <u>Elliot</u> : I'll finish making this drink and change the keg.
<i>Blame Question</i>	If you were <i>Sophie [listener]</i> how much would you blame <i>Elliot [speaker]</i> for misleading you? Rate your confidence from 0 to 5, with 0 being 'not at all' and 5 being 'completely'. [Likert scale]
<i>Mistrust Question</i>	If you were <i>Sophie [listener]</i> how much would you mistrust <i>Elliot [speaker]</i> in the future? Rate your confidence from 0 to 5, with 0 being 'not at all' and 5 being 'completely'. [Likert scale]

3.2 Results

All analyses reported here were performed using R version 3.4.1 (R Core Team 2020), RStudio (RStudio Team 2020). To test our hypotheses, we computed separate multiple linear regression models for each of our dependent variables (blame and mistrust ratings) (see e.g., Norman 2010, The unstandardized results of this approach are summarized in Table 4; standardized coefficients are reported in the text below.

Across participants and conditions, blame and mistrust responses were significantly correlated ($r = .62, p < .001$). Nonetheless, these two measures were not equally sensitive to our experimental manipulations: our analysis only explained a significant amount of variance for blame ratings. For mistrust ratings, however, the same model did not explain a statistically significant proportion of variance ($p = .283$). Therefore, we did not interpret the effects of our manipulation on this variable and only report the results of our regression analysis for blame ratings here. Figure 2 illustrates the distribution of blame ratings across experimental conditions.

Table 4 Results of the regression analysis for the blame measure of the commitment experiment. Effects significant at the 0.05 level are printed in bold. Standardized coefficients are reported in the text below

Outcome	Model	Estimate	T	p
Blame	Intercept	3.45	22.45	< 0.001
	Denial	0.512	2.3	0.020
	Strength	-0.22	-1.14	0.253
	Level of Meaning	0.495	2.86	0.005
	Denial x Strength	-0.602	-2.39	0.017
	Denial x Level of Meaning	-0.681	2.70	0.007
	Strength x Level of Meaning	0.032	0.125	0.900

We fitted a linear model (estimated using ordinary least squares) with the three-way interaction between ‘denial’, ‘strength’ and ‘level of meaning’. Since this three-way interaction was not significant, we dropped it from the model and instead used a formula only including the respective two-way interactions between our independent variables (formula: $\text{Blame} \sim \text{‘denial’} * \text{‘strength’} + \text{‘denial’} * \text{‘level of meaning’} + \text{‘level of meaning’} * \text{‘strength’}$). This model explained a statistically significant proportion of variance ($R^2=0.08$, $F(5, 381)=6.45$, $p<.001$, adj. $R^2=0.07$). The model’s intercept, corresponding to ‘denial’ = absent, ‘strength’ = strong and ‘level of meaning’ = enrichment, was at 3.45 (95% CI [3.15, 3.75], $t(381)=22.45$, $p<.001$).

Within this model, we found a significant effect of ‘denial’ (*Present*) ($\beta=0.40$, 95% CI [0.06, 0.74]; $t(381)=2.33$, $p=.020$), and of ‘level of meaning’ (*Implicature*) ($\beta=0.40$, 95% CI [0.12, 0.67]; $t(381)=2.86$, $p=.004$). Moreover, we found a significant interaction effect between ‘strength’ (*Weak*) and ‘denial’ (*Present*) ($\beta=-0.47$,

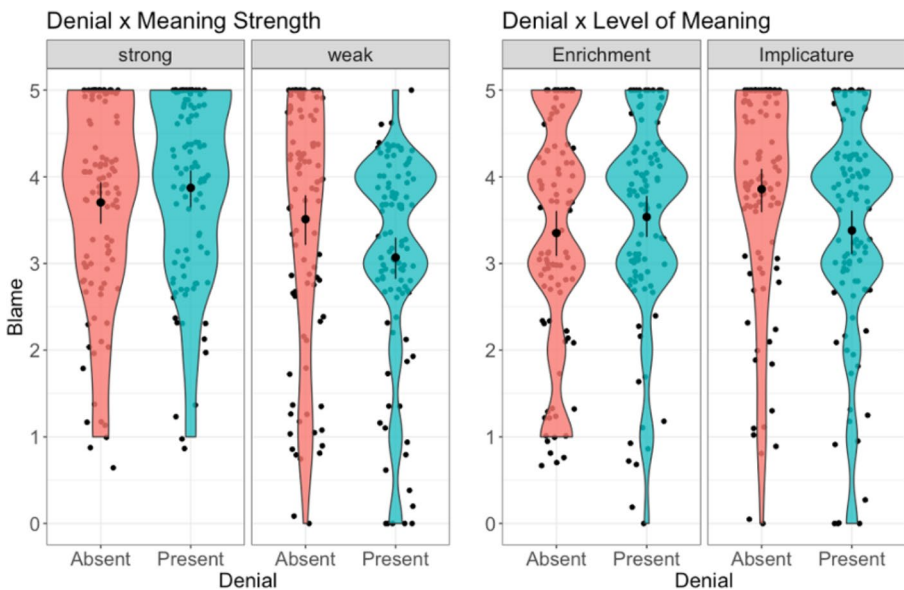


Fig. 2 Violin plots of distributions of blame ratings across meaning strengths, presence of denial, and levels of meaning. Denial was associated with lower blame ratings in Weak conditions and when meaning was conveyed via Implicature

95% CI [-0.86, -0.08]; $t(381) = -2.39, p = .017$) and a significant interaction effect between ‘level of meaning’ (*Implicature*) and ‘denial’ (*Present*) ($\beta = -0.53, 95\% \text{ CI} [-0.92, -0.15]; t(381) = -2.70, p = .007$). Standardized parameters were obtained by fitting the model on a standardized version of the dataset. These results suggest that, denial was effective at reducing blame rates in *Weak* and in *Implicature* conditions while having no effect in *Strong* and *Enrichment* conditions (see Fig. 2).

4 General Discussion

As anyone who follows daily politics knows, in order to avoid accountability people are often tempted to exploit plausible deniability: whenever a commitment is broken (a promise is not lived up to, or an assertion is found to be false), one might try to limit social repercussions by denying having committed to it in the first place. In order to refute a commitment one might deny the alleged meaning of the original utterance (‘I didn’t say that...’) or appeal to an alternative meaning (‘I meant that...’) (Boogaart et al. 2020). Pinker and colleagues have argued that speakers in certain circumstances capitalise on these options when they generate their messages (Pinker et al. 2008; Lee and Pinker 2010). In this study, we investigated what factors influence plausible deniability. What determines whether a speaker can get out of conversational commitments by denying they intended the alleged meaning of their utterance? If a commitment is plausibly deniable the costs the speaker incurs following its breach will be mitigated after denial. However, the implicit/explicit dichotomy is not the only factor at play. Our design therefore explored the impact of two factors: the degree of explicitness (the level of meaning) of what is communicated and its strength.

Our findings indicate that both meaning strength and level of meaning influence plausible deniability: in the presence of a denial, participants judged the speaker to be less blameworthy when the implied content was weakly conveyed (but not when it was strongly conveyed), or when this content was an implicature (but not an enrichment). In view of the results of the norming study it is unlikely that these observations could be attributed to participants failing to draw the intended implicatures from our materials.

As predicted, our results show that denial works better in weak contexts than in strong contexts: weak contents showed higher deniability (i.e., they were associated with lower blame ratings after denial) than strong contents, for which blame rates remained unaffected by denial. This is one of the key findings of our study. The fact that weakly conveyed contents are easier to deny is consistent with Nicolle and Clark (1999), who found that strongly conveyed meanings (in contrast to weakly conveyed ones) are frequently associated with ‘what is said’. Our results also reinforce the conclusions of Sternau et al. (2015): when explicitly asked, their participants found implicatures more deniable than enrichments, and weak implicatures more deniable than strong implicatures. They are also consistent with the considerations by Dinges and Zakkou (2023), who argue that deniability is an epistemic feature preventing an audience to reason or form knowledge about what was communicated. Together, these findings confirm that, as argued by Mazzarella (2021), meaning strength exerts a considerable influence on plausible deniability.

Importantly, the current study also shows that levels of meaning modulate plausible deniability. Implicatures appeared to be more easily deniable (with blame rates lower after denial) than enrichments (blame rates were unaffected by denial). This is reminiscent of Reins and Wiegmann (2021), who found a similar pattern relying on participants' *a priori* intuitions about denial. Our findings also echo those of Hall and Mazzarella (2023) who found participants would trust more a speaker uttering a false enrichment than one conveying a false implicature. We used four different types of enrichments – conjunction enrichment, conditional perfection, as well as scalar inferences linked to the quantifier *some*, and to disjunction; our results are therefore not reduceable to a single pragmatic phenomenon, instead they can be, to some extent, generalised to enrichments in general (see also, Hall and Mazzarella 2023, for how different types of enrichments might affect speaker accountability). One caveat worth noting, however, is that – since scenarios were not matched for content – we cannot currently rule out that the effect of levels of meaning was driven to some extent by differences in content.

These findings are also consistent with the contextualist claim that enrichments can be included into the truth-conditions of the utterance (i.e., 'what is said') (Récanati 1993, 2001, 2004; Carston 2002; Sperber and Wilson 1986/95). If enrichments affect the truth-conditional content of the utterance and are perceived as being part of 'what is said', then they should, *ceteris paribus*, be harder to deny. Our results are thus in line with the those of Doran et al. (2012) indicating that, contrary to implicatures, enrichments are sometimes included in 'what is said'; Weissman and Terkourafi (2018), who show that false enrichments are considered lies, while false implicatures are not (see below for further discussion); as well as Bonalumi et al. (2020), who highlight that unfulfilled enriched promises are considered *de facto* broken promises, while unfulfilled implicated promises are not.

We collected two measures of accountability: a blameworthiness judgment and an untrustworthiness judgment. While the two judgments were correlated, we found no evidence that mistrust judgments were affected by our experimental manipulations. One possible reason for this discrepancy might be that the mistrust question was always presented after the blame question. It may be that some participants, after having simulated a partner-control strategy (i.e., blaming the speaker for falsely implying something), perceived an additional partner-choice strategy (i.e., mistrusting the speaker) as redundant or excessively severe (see Noë and Hammerstein 1994). It may also be the case that blameworthiness is more severely impacted by pragmatic factors than trustworthiness. It could also be that the different utterances presented in the stimuli were themselves carrying disparate illocutionary force, causing an array of other expectations that were instead (partially) satisfied. There would then be no reason left to consider the speaker untrustworthy, despite still being blameworthy for the misleading utterance. Finally, it may be that the phrasing of the mistrust question was confusing for the participants – leading them to interpret it as a *trust* question.

Interestingly, some very recent studies have linked the concept of lying to commitment (Reins and Wiegmann 2021; Wiegmann et al. 2021). Specifically, they argue that participants' intuitive understanding of lying does not only encompass intentionally false assertions (part of 'what is said'), but also false implicatures to which the speaker seems committed (against traditional views, which restrict lying to what is

explicitly asserted, see, e.g., Carson 2006, 2010; Saul 2012; Stokke 2018).⁶ Claims concerning the nature of lying and whether it should include misleading implicit communication are beyond the scope of our work. Nevertheless, our findings have interesting consequences for views which equate lying with intentionally communicated misleading content, including implicit content – as long as the speaker appears sufficiently committed to it. If commitment is central to an analysis of lying, it is essential to investigate markers of commitment. From this perspective, the present study precisely highlights two factors routinely used to modulate commitment: the speaker’s choice of level of meaning and the strength of the implicit content. Indeed, the difference we find between enrichments and implicatures, both in terms of accountability and of deniability, echoes the results of Reins and Wiegmann (2021). Even more crucially, our findings emphasize the role of meaning strength in how committed the speaker appears to be to a given implicit content: meaning strength seems to play a crucial role in shaping our intuitions about commitment. If speaker commitment is the key to the hearer’s perception of truth and lying – and their related social consequences – it is critical to further investigate the role of meaning strength.

When speakers aim to limit their commitment and the social consequences of communicating potentially unwelcome or unreliable messages, one good strategy is to increase the room for plausible deniability (Lee and Pinker 2010; Pinker 2007; Pinker et al. 2008). The results presented here provide clear evidence of the effective impact of a denial on speaker’s accountability. Our findings suggest that accountability can only be mitigated by denial when certain pragmatic factors are in place: despite being logically cancellable, not all contents conveyed implicitly by the speaker can be plausibly denied. This will depend on the degree of explicitness and the strength with which the message is communicated. Only when these are properly aligned can the speakers optimally minimise the social consequences of an unreliable message. Strategic communication is a complex phenomenon that goes beyond the implicit/explicit dichotomy.

Acknowledgements We would like to thank Péter Rácz for his help with statistical analyses, as well as Richard Breheny for valuable advice and support throughout the project, and Keith Jensen and Christophe Heintz for the helpful comments and discussion.

Authors’ contributions *Francesca Bonalumi* Conceptualisation, Methodology, Writing - Original Draft, Writing - Review & Editing, Visualisation. *Johannes B. Mahr* Conceptualisation, Methodology, Formal analysis, Data curation, Writing - Review & Editing. *Pauline Marie* Conceptualisation, Methodology, Software, Investigation. *Nausicaa Pouscoulous* Conceptualisation, Methodology, Resources, Writing - Original Draft, Writing - Review & Editing, Supervision, Project administration.

Funding Open access funding provided by Central European University Private University

⁶ As we mention in the Introduction, in some studies participants do not consider most deceptive implicatures as lies (Weissman and Terkourafi 2019), while they appear to in others (Antomo et al. 2018; Or et al. 2017; Reins and Wiegmann 2021; Wiegmann et al. 2021; Wiegmann & Willemsen, 2017).

Declarations

Ethical Approval All data reported in this manuscript were collected in accordance with the guidelines set out by the UCL ethical review board and the experiment was approved by the Ethics Chair of the UCL Department of Linguistics.

Conflict of Interest The author(s) declared no conflicts of interest with respect to the authorship or the publication of this article.

Open Practices Statement The Commitment Experiment reported in this article was formally pre-registered on OSF.io. The pre-registration document is available at: <https://osf.io/nkv63/>. The data and the materials of both experiments are available at: <https://osf.io/2fu93/>.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Antomo, M., S. Müller, K. Paul, M. Paluch, and M. Thalmann. 2018. When children aren't more logical than adults: an empirical investigation of lying by falsely implicating. *Journal of Pragmatics* 138: 135–148. <https://doi.org/10.1016/j.pragma.2018.09.010>.
- Beysade, C., and J.-M. Marandin. 2009. Commitment: une attitude dialogique. *Langue Française* 2: 89–197. <https://doi.org/10.3917/lf.162.0089>.
- Bezuidenhout, A., and J. C. Cutting. 2002. Literal meaning, minimal propositions, and pragmatic processing. *Journal of Pragmatics* 34(4): 433–456. [https://doi.org/10.1016/S0378-2166\(01\)00042-X](https://doi.org/10.1016/S0378-2166(01)00042-X).
- Bonalumi, F., T. Scott-Phillips, J. Tacha, and C. Heintz. 2020. Commitment and communication: are we committed to what we mean, or what we say? *Language and Cognition* 1–25. <https://doi.org/10.1017/langcog.2020.2>.
- Boogaart, R., H. Jansen, and M. van Leeuwen. 2020. Those are your words, not mine! Defence strategies for denying Speaker Commitment. *Argumentation* 35: 209–235. <https://doi.org/10.1007/s10503-020-09521-3>.
- Boulat, K., and D. Maillat. 2017. She said you said I saw it with my own eyes: a pragmatic account of commitment. In *Formal models in the study of Language*, eds. J. Blochowiak, C. Grisot, S. Durrleman, and C. Laenzlinger. 261e279. Cham: Springer.
- Boulat, K., and D. Maillat. 2023. Strength is relevant: Experimental evidence of strength as a marker of commitment. *Frontiers in Communication*, 8. <https://doi.org/10.3389/fcomm.2023.1176845>.
- Brown, P., and S. C. Levinson. 1987. *Politeness: some universals in language usage*. Cambridge University Press.
- Camp, E. 2018. Insinuation, Common Ground, and the conversational record. In *New Work on Speech Acts*, eds. D. Fogal, D. W. Harris, and M. Moss. vol. 1 40–66. Oxford University Press. <https://doi.org/10.1093/oso/9780198738831.003.0002>.
- Camp, E. 2022. Just saying, just kidding: Liability for accountability-avoiding speech in ordinary conversation, politics and law. In L. R. Horn (Ed.), *From Lying to Perjury* (pp. 227–258). De Gruyter Mouton. <https://doi.org/10.1515/9783110733730-010>.
- Carson, T. L. 2006. The definition of lying. *Noûs* 40(2): 284–306. <https://doi.org/10.1111/j.0029-4624.2006.00610.x>.

- Carson, T. L. 2010. *Lying and deception: theory and practice*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199577415.001.0001>.
- Carston, R. 2002. *Thoughts and utterances: the Pragmatics of Explicit Communication*. Blackwell.
- Clark, B. 2013. *Relevance theory*. Oxford University Press.
- Danziger, E. 2010. On trying and lying: Cultural configurations of Grice's Maxim of Quality. *Intercultural Pragmatics* 7(2): 199–219. <https://doi.org/10.1515/iprg.2010.010>.
- De Brabanter, P., and P. Dendale. 2008. Commitment: the term and the notions. *Belgian Journal of Linguistics* 22: 1–14. <https://doi.org/10.1075/bjl.22.01de>.
- Dinges, A., and J. Zakkou. 2023. *On Deniability Mind* 132(526): 372–401. <https://doi.org/10.1093/mind/fzac056>.
- Doran, R., G. Ward, M. Larson, Y. McNabb, and R. E. Baker. 2012. A novel experimental paradigm for distinguishing between what is Said and what is implicated. *Language* 88(1): 124–154. <https://doi.org/10.1353/lan.2012.0008>.
- Doran, R., R. Baker, Y. McNabb, M. Larson, and G. Ward. 2009. On the Non-Unified Nature of Scalar Implicature: an empirical investigation. *International Review of Pragmatics* 1: 1–38. <https://doi.org/10.1163/187730909X12538045489854>.
- Faul, F., E. Erdfelder, A. Buchner, and A.-G. Lang. 2009. Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses. *Behavior Research Methods* 41(4): 1149–1160. <https://doi.org/10.3758/BRM.41.4.1149>.
- Franke, M., G. Dulcinati, and N. Pouscoulous. 2020. Strategies of deception: Under-Informativity, Uninformativity, and lies-misleading with different kinds of implicature. *Topics in Cognitive Science* 12(2): 583–607. <https://doi.org/10.1111/tops.12456>.
- Geurts, B. 2019. Communication as commitment making: Speech acts, implicatures, common ground. *Theoretical Linguistics* 45(1–2): 1–30. <https://doi.org/10.1515/tl-2019-0001>.
- Gibbs, R. W., and J. F. Moise. 1997. Pragmatics in understanding what is said. *Cognition* 62(1): 51–74. [https://doi.org/10.1016/S0010-0277\(96\)00724-X](https://doi.org/10.1016/S0010-0277(96)00724-X).
- Grice, H. P. 1989. *Studies in the way of words*. Harvard University Press.
- Hall, A., and D. Mazzarella. 2023. Pragmatic inference, levels of meaning and speaker accountability. *Journal of Pragmatics* 205: 92–110. <https://doi.org/10.1016/j.pragma.2022.12.007>.
- Hamblin, C. L. 1971. Mathematical models of dialogue 1. *Theoria*, 37(2).
- Hardin, K. J. 2010. The spanish notion of lie: Revisiting Coleman and Kay. *Journal of Pragmatics* 42(12): 3199–3213. <https://doi.org/10.1016/j.pragma.2010.07.006>.
- Horn, L. 1989. *A natural history of negation*. Chicago University Press.
- H.R. 5430–116th Congress: United States-Mexico-Canada Agreement Implementation Act. 2022. Retrieved from <https://www.govtrack.us/congress/bills/116/hr5430>.
- Hruschka, D. 2020. *Cultural Diversity in the Meaning of Lies, Deceptions, and other Misrepresentations*. SocArXiv. <https://doi.org/10.31235/osf.io/8puwc>.
- Lee, J. J., and S. Pinker. 2010. Rationales for Indirect Speech: the theory of the Strategic Speaker. *Psychological Review* 117(3): 785–807. <https://doi.org/10.1037/a0019688>.
- Levinson, S. C. 2000. *Presumptive meanings. The theory of generalized conversational implicature*. The MIT Press.
- Mahr, J. B., and G. Csibra. 2018. Why do we remember? The communicative function of episodic memory. *Behavioral and Brain Sciences* 41: 1–16. <https://doi.org/10.1017/S0140525X17000012>.
- Mahr, J. B., and G. Csibra. 2020. Witnessing, remembering, and testifying: why the past is special for human beings. *Perspectives on Psychological Science* 15(2): 428–443. <https://doi.org/10.1177/1745691619879167>.
- Mahr, J. B., and G. Csibra. 2021. The effect of source claims on statement believability and speaker accountability. *Memory & Cognition* 49(8): 1505–1525. <https://doi.org/10.3758/s13421-021-01186-x>.
- Mazzarella, D. 2021. I didn't mean to suggest anything like that!: deniability and context reconstruction. *Mind & Language* 1–19. <https://doi.org/10.1111/mila.12377>.
- Mazzarella, D., R. Reinecke, I. Noveck, and H. Mercier. 2018. Saying, presupposing and implicating: how pragmatics modulates commitment. *Journal of Pragmatics* 133(August): 15–27. <https://doi.org/10.1016/j.pragma.2018.05.009>.
- Morency, P., S. Oswald, and L. De Saussure. 2008. Explicitness, implicitness and commitment attribution: a Cognitive Pragmatic Approach. *Belgian Journal of Linguistics* 22(1): 197–219. <https://doi.org/10.1075/bjl.22.10mor>.
- Nicolle, S., and B. Clark. 1999. Experimental pragmatics and what is said: a response to Gibbs and Moise. *Cognition* 69(3): 337–354. [https://doi.org/10.1016/S0010-0277\(98\)00070-5](https://doi.org/10.1016/S0010-0277(98)00070-5).

- Noë, R., and P. Hammerstein. 1994. Biological Markets: supply and demand determine the Effect of Partner Choice in Cooperation, Mutualism and mating. *Behavioral Ecology and Sociobiology* 35(1): 1–11.
- Norman, G. 2010. Likert scales, levels of measurement and the laws of statistics. *Adv in Health Sci Educ* 15: 625–632. <https://doi.org/10.1007/s10459-010-9222-y>.
- Noveck, I. A. 2004. Pragmatic Inferences Related to Logical Terms. In I. A. Noveck & D. Sperber (Eds.), *Experimental Pragmatics. Palgrave Studies in Pragmatics, Language and Cognition* (pp. 301–321). Palgrave Macmillan UK. https://doi.org/10.1057/9780230524125_14.
- Noveck, I. A., and D. Sperber. 2007. The why and how of experimental pragmatics: the case of ‘scalar inferences’. In *Advances in Pragmatics*, 184–212. Palgrave.
- Or, S., M. Ariel, and O. Peleg. 2017. The case of literally true propositions with false implicatures. In *Deception & deceptive communication. Motivations, Recognition techniques and behavioral control*, ed. I. Chilawa. Nova Science.
- Palan, S., and C. Schitter. 2018. Prolific.ac – a subject pool for online experiments. *Journal of Behavioral and Experimental Finance* 17: 22–27. <https://doi.org/10.1016/j.jbef.2017.12.004>.
- Pinker, S. 2007. The evolutionary social psychology of off-record indirect speech acts. *Intercultural Pragmatics* 4(4): 437–461. <https://doi.org/10.1515/IP.2007.023>.
- Pinker, S., M. A. Nowak, and J. J. Lee. 2008. The logic of indirect speech. *Proceedings of the National Academy of Sciences*, 105(3), 833–838. <https://doi.org/10.1073/pnas.0707192105>.
- Récánati, F. 1993. *Direct reference: from Language to Thought*. Blackwell.
- Récánati, F. 2001. What is Said. *Synthese* 128(1–2): 75–91. <https://doi.org/10.1023/A:1010383405105>.
- Récánati, F. 2004. *Literal meaning*. Cambridge University Press.
- R Core Team. 2020. *R: A language and environment for statistical computing. R Foundation for Statistical Computing*. <https://www.R-project.org/>.
- Reboul, A. 2017. Is implicit communication a way to escape epistemic vigilance? In *Pragmatics at its interfaces*, ed. S. Assimakopoulos. 91–112. de Gruyter Mouton.
- Reins, L. M., and A. Wiegmann. 2021. Is lying bound to commitment? Empirically investigating deceptive Presuppositions, Implicatures, and actions. *Cognitive Science* 45(2): e12936. <https://doi.org/10.1111/cogs.12936>.
- RStudio: Integrated Development for R*. RStudio Team, and P. B. C. RStudio. 2020. <http://www.rstudio.com/>.
- Saul, J. M. 2012. *Lying, Misleading, and what is Said*. Oxford University Press.
- Soltys, J., M. Terkourafi, and N. Katsos. 2014. Disentangling politeness theory and the Strategic Speaker approach: theoretical considerations and empirical predictions. *Intercultural Pragmatics* 11(1): 31–56. <https://doi.org/10.1515/ip-2014-0002>.
- Sperber, D., and D. Wilson. 1986. *Relevance: communication and cognition*. Oxford: Blackwell Publishing.
- Sperber, D., and D. Wilson. 2006. Pragmatics. In *Oxford Handbook of Philosophy of Language*, eds. E. Lepore, and B. C. Smith. Oxford University Press.
- Starmans, C., and O. Friedman. 2012. The folk conception of knowledge. *Cognition* 124(3): 272–283. <https://doi.org/10.1016/j.cognition.2012.05.017>.
- Sternau, M., M. Ariel, and R. Giora. 2017. Deniability and explicatures. In *Doing Pragmatics intercultural*, eds. R. Giora, and M. Haugh. 97–120. De Gruyter Mouton.
- Sternau, M., M. Ariel, R. Giora, and O. Fein. 2015. Levels of interpretation: New tools for characterizing intended meanings. *Journal of Pragmatics* 84: 86–101. <https://doi.org/10.1016/j.pragma.2015.05.002>.
- Stokke, A. 2018. *Lying and insincerity*. Oxford University Press. <https://doi.org/10.1093/oso/9780198825968.001.0001>.
- Viebahn, E., A. Wiegmann, and P. Willemsen. 2018. *Can a question be a lie? An empirical investigation*.
- Vullioud, C., F. Clément, T. Scott-Phillips, and H. Mercier. 2017. Confidence as an expression of commitment: why misplaced expressions of confidence backfire. *Evolution and Human Behavior* 38(1): 9–17. <https://doi.org/10.1016/j.evolhumbehav.2016.06.002>.
- Weissman, B., and M. Terkourafi. 2019. Are false implicatures lies? An empirical investigation. *Mind & Language* 34(2): 221–246. <https://doi.org/10.1111/mila.12212>.
- Wiegmann, A. 2022. Lying with deceptive implicatures? Solving a puzzle about conflicting results. *Analysis*. anac037.
- Wiegmann, A., P. Willemsen, and A. Wiegmann. 2017. How the truth can make a great lie: An empirical investigation of the folk concept of lying by falsely implicating. *Proceedings of the 39th Annual Conference of the Cognitive Science Society*, 3516–3621.

- Wiegmann, A., P. Willemsen, and J. Meibauer. 2021. *Lying, Deceptive Implicatures, and Commitment* [Preprint]. PsyArXiv. <https://doi.org/10.31234/osf.io/n96eb>.
- Wilson, D., and D. Sperber. 2002. Truthfulness and relevance. *Mind* 111(443): 583–632.
- Wilson, D., and D. Sperber. 2004. Relevance theory. In *Handbook of Pragmatics*, eds. L. Horn, and G. Ward. 607–632. John Wiley & Sons, Ltd.
- Yuan, W., and S. Lyu. 2022. Speech act matters: commitment to what's said or what's implicated differs in the case of assertion and promise. *Journal of Pragmatics* 191: 128–142. <https://doi.org/10.1016/j.pragma.2022.01.012>.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Francesca Bonalumi¹ · Johannes B. Mahr² · Pauline Marie³ ·
Nausicaa Pouscoulous³

-
- ✉ Francesca Bonalumi
bonalumif@ceu.edu
 - ✉ Johannes B. Mahr
jmahr@yorku.ca
 - ✉ Pauline Marie
paulinemarie8@gmail.com
 - ✉ Nausicaa Pouscoulous
n.pouscoulous@ucl.ac.uk

¹ Department of Cognitive Science, Central European University, Quellenstraße 51, Vienna 1100, Austria

² Department of Philosophy, York University, Toronto, Canada

³ Department of Linguistics, University College London, London, UK