Cognition 237 (2023) 105463

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

Pragmatic responses to under-informative *some*-statements are not scalar implicatures☆

Mikhail Kissine [a,b,c,*], Philippe De Brabanter [b]

[a] *ACTE, Université libre de Bruxelles, Belgium*
[b] *LaDisco, Université libre de Bruxelles, Belgium*
[c] *ULB Neuroscience Institute, Université libre de Bruxelles, Belgium*

ARTICLE INFO

Dataset link: https://osf.io/c78sd

Keywords:
Scalar implicatures
Truth-judgement task
Linguistic representation
Grammatical derivation

ABSTRACT

A highly emblematic paradigm in experimental pragmatics consists in presenting participants with an existentially quantified sentence of the form *Some X are Y* in a context in which all X are obviously Y. Participants who reject such sentences as false or infelicitous are said to adopt a 'pragmatic' instead of a 'logical' reading of *some*, and to derive the scalar implicature *Some, but not all X are Y*. Although there are several competing accounts of scalar implicatures, virtually all of them assume that a participant who responds pragmatically to an under-informative *some*-sentence mentally entertains a linguistic representation of the negation of a stronger alternative (*All X are Y*). Yet, there is no evidence that judging an under-informative *some*-sentence false or infelicitous actually involves the derivation of the *some, but not all* scalar implicature. We report three experiments consisting of a sentence-picture verification task followed by a forced choice between two paraphrases of the sentence initially assessed. These experiments robustly show that hearers who reject an under-informative *some*-sentence do so without explicitly entertaining a *some, but not all* implicature. Our results represent a strong challenge for grammatical accounts of scalar implicature, which all presuppose a mechanism of negation of stronger alternatives, and force a drastic reinterpretation of processing data on scalar implicatures. More generally, our findings show that one should not conflate psychological models of pragmatic processing with a reconstructed link between sentences and their potential meanings.

1. Introduction

Ira Noveck's seminal (2001) paper prompted an incredibly rich line of inquiry into the nature of the link between quantified sentences like (1a) and the scalar implicatures, such as (1b), that their utterance is generally expected to trigger.

(1) a. The basketball player scored some of his shots.
b. ⇝ The basketball player scored some, but not all of his shots.
c. The basketball player scored all of his shots.

The link between (1a) and (1b) has traditionally been modelled in broadly Gricean terms. Rehashing the steps of such a Gricean reconstruction would be tedious for most readers (see, for instance, Geurts, 2010, for a pedagogical introduction). A point worth recalling, however, is that in Grice's explicit aim was to provide a rational reconstruction of speaker behaviour, that is, to model it using premises and inferential steps that do not stand at odds with broadly accepted principles of rationality (Grice, 1989). In the case of (1a), such a reconstruction may look like this:

1. S can be assumed to strive to be the most informative possible given the purposes of the conversation and to avoid asserting things for which they do not have evidence [Gricean principles of Cooperation]
2. (1a) is entailed by the stronger alternative (1c)
3. It would not be rational for S to utter (1a) if S believed that the basketball player scored all of his shots [from 1 and 2]
4. S is competent about the shots the player scored. [Competence assumption]
5. S believes that the player did not score all of his shots [from 3 and 4][1]

☆ This paper is a part of special issue "20 Years of XPrag".

* Correspondence to: CP 175, 50 av. F.D. Roosevelt, 1050 Bruxelles, Belgium.
E-mail address: Mikhail.Kissine@ulb.be (M. Kissine).

[1] It is, of course, possible that the Competence assumption does not hold, in which case the implicature derived is the weaker *S does not believe that the player scored all of his shots*. However, the distinction between weak and strong implicatures does not play any decisive role in what follows, and we will set it aside from now on.

https://doi.org/10.1016/j.cognition.2023.105463
Received 25 July 2022; Received in revised form 2 April 2023; Accepted 5 April 2023
Available online 13 April 2023
0010-0277/

∴ By uttering (1a), S conveys that (1b)

While the reconstruction just sketched reflects the speaker's perspective, most of the experimental pragmatic research centres on the interpreter's side (but see, e.g. Benz & Gotzner, 2014, 2020; van Tiel et al., 2021). Crucially, in transposing the Gricean rational reconstruction to the interpreter's side, the determining step for the derivation of the scalar implicature is 2. Knowing that the speaker had a stronger alternative available, the Gricean hearer has to decide whether the contextual evidence suffices to dismiss this stronger alternative.[2] Virtually all the experimental research assumes, under one guise or another, that, in the process of pragmatically interpreting *some*, hearers rely on the negation of the stronger alternative. In this paper, however, we question whether this central assumption holds, at least for paradigmatic test cases. Our reason for doing so is that, as we will argue now, this issue has crucial consequences for adjudicating between different accounts of scalar implicatures, but also, more generally, for the way results from truth- or felicity-judgement tasks translate to psychological models of pragmatic processing.

1.1. 'Blind' implicatures

In the wake of the publication of Noveck (2001), the existential quantifier became the 'drosophila of the experimental pragmatic literature' (Noveck, 2018, 79), and scalar implicatures emerged as the central concern of that new field of research. One key finding of Noveck (2001, Exp. 3) is that adult participants tend to judge sentences like (2) false, even though these sentences are logically true.[3]

(2) Some giraffes have long necks.

The almost universally favoured interpretation of this result is that a participant who judges (2) false does so because they interpret *some* 'pragmatically', as *some, but not all*. It is remarkable (and duly noted by, e.g., Chemla & Singh, 2014b; Magri, 2017; Schlenker, 2012) that, although cases like (2) sparked a whole new way of doing pragmatics, they are not amenable to classic Gricean reconstructions, because the context makes it impossible for a rational hearer to judge the stronger alternative as false. It is indisputable that examples such as (2) do sound odd. And it does seem plausible that this oddity can be traced back to the fact that the pragmatically reinforced meaning *some, but not all* contradicts common knowledge. Recall, however, that from a Gricean point of view, whether or not a scalar implicature is derived depends on whether there are sufficient contextual grounds for negating the stronger alternative. In cases like (2), the context should preclude the negation of the stronger alternative, because that negation is obviously false. Accordingly, if (2) gives rise to the derivation of a scalar implicature, then this derivation should be 'blind' to contextual knowledge (Magri, 2017; Schlenker, 2012). A corollary, then, would be that *some* has a privileged status, as it triggers the pragmatically strengthened reading even in a context where this reading is manifestly false. This also means that the relevant set of alternatives should be specified in a context-independent way, which is not a trivial task from a formal point of view (see Magri, 2017; Schlenker, 2012).

It has been objected (Guasti et al., 2005) that participants may try to find an interpretation on which (2) is unambiguously true (for instance, by thinking of baby giraffes, which have comparatively shorter necks). As an alternative to experimental paradigms that rely on world knowledge, many studies turned to truth- or accuracy-judgement tasks, in which participants have to decide whether the target sentence matches a visual display. The idea here, just as in Noveck's (2001) original task, is to prompt the derivation of implicatures blind to the context. For instance, participants who judge that (3) is not a good description of a picture in which all monkeys are eating a biscuit, in Guasti et al. (2005), or those who judge (4) to be false in a situation where all shapes are red, in van Tiel and Schaeken (2017), are taken to have derived the corresponding *some, but not all* interpretation even though the experimental context makes it manifestly impossible to judge the stronger alternative false:

(3) Some monkeys are eating a biscuit.
(4) Some of the shapes are red.

1.2. Explicit Derivation Hypothesis

Such blind implicatures are also an essential source of evidence in the literature on pragmatic processing. The impetus here was given by Bott and Noveck (2004), who reported that it takes longer to judge a stimulus like (2) false than it takes to judge it true. It is not this precise result that is important here, but the fact that in this and the subsequent studies, the time it takes to judge an under-informative *some*-sentence false is understood to mirror an actual aspect of cognitive processing—the derivation of the corresponding scalar implicature. Implicitly or not, these experimental studies thus share the assumption that, from a processing point of view, scalar implicatures correspond to the derivation of one linguistic representation, *some, but not all X are Y*, from another, *some X are Y* (see Breheny, 2019, for a recent review). This assumption entails the following empirical hypothesis about the processing of blind scalar implicatures:

> **Explicit Derivation Hypothesis**
>
> Judging a sentence of the form *Some X are Y* to be false in a situation where it is obvious that all X are Y entails explicitly representing the reinforced reading *Some, but not all X are Y*[4]

We will now discuss two important points in connection with this **Explicit Derivation Hypothesis**. First, it is held more or less across the board in experimental pragmatics, i.e. independently of theoretical orientation. Second, and rather surprisingly, no empirical evidence has been adduced in its favour.

According to 'grammatical' models, the derivation of scalar implicatures primarily belongs to syntactic computation. Such models may be couched in terms of the narrowing of the lexical meaning of *some* (Potts et al., 2015) or of the differential scope of a covert exhaustivity operator (e.g. Cherchia, 2004; Fox & Katzir, 2011), but regardless of the exact implementation, they all endorse the **Explicit Derivation Hypothesis**. Chemla and Singh (2014a, 2014b) argue that performance properties should map onto (grammatical and pragmatic) competence, which encompasses the computational mechanism allowing the derivation of one meaning of *some* from the other. Surprisingly, though, the proponents of grammatical accounts of scalar implicatures (or, for that matter, those of formal semantic models) do not provide evidence that in the critical cases the *some, but not all* string is actually derived. Bott and Chemla (2016) do show that when a 'pragmatic' interpretation is forced in one experimental trial, this increases the likelihood that the same kind of interpretation will be spontaneously chosen in the

[2] The relationship between *some* and the stronger alternative *all* is that of entailment. However, as pointed out by an anonymous reviewer, scalar implicatures do not require such a genuine entailment relationship; for instance, the use of *cheap* may give rise to the implicature not-*free*, even though it is not absolutely obvious that *free* entails *cheap*.

[3] Noveck (2001) also found that children tended to provide more logical, viz. true, judgements to those kinds of sentences. This result sparked exciting and very fruitful research in developmental pragmatics (e.g. Papafragou & Skordos, 2016). Here, however, we will limit ourselves to adult participants, because we are concerned with a very basic and simple-minded question: What does it mean for a participant to judge (2) false?

[4] Or, equivalently, explicitly representing the negation of the stronger alternative *All X are Y*, which, in conjunction with *Some X are Y*, yields *Some, but not all X are Y*.

subsequent trial. However, this constitutes evidence that it is possible to prime a pragmatic 'strategy', not that the actual implicatures are represented.

Moving to the neo-Gricean camp, where scalar implicatures are seen as post-hoc pragmatic inferences, Geurts and Pouscoulous (2009, Exp. 2) compared participants' reactions to under-informative sentences such as (5) in two conditions. In the first 'inference' condition participants were explicitly asked about the implicature the target sentence (putatively) generates, e.g. whether the sentence implied that some of the Bs were not in the box on the left; in the second 'verification' condition the same participants had to decide whether the sentence was true or false.

(5) Some of the Bs are in the box on the left.
Situation: [BBBAAA] [CCC]

Interestingly, the rates of 'pragmatic' responses in the verification condition were significantly lower (34%) than in the inference condition (62%). In Geurts and Pouscoulous (2009), the explicit choice of the implicature in the inference condition and the response 'false' in the verification condition are placed on the same footing. Geurts and Pouscoulous (2009) assume both to be diagnostic of the derivation of the implicature *Some, but not all of the Bs are in the box on the left* by the respondent. However, they propose that explicitly asking whether a sentence is associated with an implicature biases respondents towards deriving it. In fact, Geurts and Pouscoulous (2009, Exp. 3) invoke their results to justify the use of this verification method to test the existence of 'embedded implicatures' (a matter to which we will return in the General Discussion).

For other pragmatic theories of scalar implicatures, the processing properties underlying participant assessment of examples like (2) correspond to local context-driven interpretation processes, which yield the enriched lexical meaning *some, but not all*. For instance, Papafragou and Musolino (2003, Exp. 1) had adult and child participants judge the felicity of under-informative *some* sentences uttered by a puppet; in this task adults almost never judged such under-informative sentences felicitous (i.e. reported that the puppet 'answered well') and they justified their responses by invoking the strongest scale-mate *all*. The authors interpret their results as demonstrating the 'psychological reality of scalar implicatures' (Papafragou & Musolino, 2003, 267-8; see also Guasti et al., 2005). That is, just like grammatical models, pragmatic theories of scalar implicature appear to endorse the **Explicit Derivation Hypothesis**.

1.3. No-Implicature Hypothesis

One exception to this trend is Katsos and Bishop (2011), who essentially replicate the results of Papafragou and Musolino (2003, Exp. 1), but argue that in order to judge *Some X are Y* infelicitous or false in a context where all X are Y, participants need not derive the actual *Some but not all X are Y* implicature, but simply have to be aware 'that a more informative statement could have been made' (Katsos & Bishop, 2011, 69). The idea is that participants who decide that under-informative but logically true *some*-sentences are false or contextually inappropriate do so without actually deriving the meaning *some, but not all* (see also Katsos, 2009). Although Katsos and Bishop (2011) provide no empirical confirmation for this hypothesis, it is interesting to note that in their study, and also in Papafragou and Musolino (2003, Exp. 1), the justifications provided by adult participants for rejecting the under-informative *Some X are Y* stimuli referred to the fact that all X were Y in the situation they were presented with, rather than mentioning the reinforced *some, but not all* meaning. Obviously, though, such justifications offer no clear indication as to whether actual derivation of the scalar implicature *Some but not X are Y* took place or not.

Although Noveck (2018, 89-90) explicitly rejects Katsos and Bishop's (2011) interpretation of the investigation of scalar implicature in a truth-judgement task, his work is framed within Relevance theory, which is fully compatible with the idea that no explicit scalar implicature is actually derived in these cases. Relevance theory draws a distinction between 'explicatures', defined as pragmatic enrichments of the linguistic meaning of a sentence, and 'implicatures', understood as pragmatically inferred propositions that are independent of the (enriched) sentence meaning (e.g. Carston, 2002; Sperber & Wilson, 1995). It is pretty much uncontroversial that, in Relevance-theoretic parlance, (1b) should be an enrichment of (1a), and hence not an implicature. Minor as this terminological point may seem, it does raise an important issue beyond the Relevance-theoretic framework. In standard Relevance theory, the derivation of explicatures and implicatures is posited to unfold on-line, the former warranting the derivation of the latter (e.g. Carston, 2002). However, as argued by Jary (2013, 2016), in many cases, the derivation of the explicature does not constitute a necessary inferential step towards the contextually relevant implicature. This is certainly the case for 'blind' scalar implicatures. The experimental setting and instructions inevitably make the decision whether the stimulus sentence is true or false (or felicitous or not) the most relevant aspect of its interpretation. Rewording the gist of Grice's idea about informativeness, at one level or another, we may also assume that adult participants are experienced communicators who have implicitly internalised that in a situation where all X are Y a cooperative speaker would usually say *All X are Y* rather than *Some X are Y*. We may also assume that a Gricean hearer would find false utterances uncooperative. But then, judging (1a) to be false in a situation where the player scored all his shots may be reconstructed as follows:

1. The player scored all his shots [Visual context]
2. This situation is described as (1a) [Linguistic stimulus]
3. Usually a cooperative speaker would not use *some* in this situation [Gricean assumption (Quantity)]
4. It is not appropriate to use (1a) in this situation [from 1, 2 and 3]
5. If a sentence is false, then it is contextually inappropriate [Gricean assumption (Quality)]

⇝ (1a) is false [Affirming the consequent, from 4 and 5]

Three important points are worth discussing at this stage. First, nowhere in this reconstruction do we need to derive the actual scalar implicature (1b) from (1a) to reach the conclusion that (1a) is conversationally inadequate. Accordingly, someone who seeks to provide a pragmatic explanation of 'blind' implicatures does not have to posit the derivation of the actual implicature *some, but not all*, resulting from the negation of the stronger alternative with *all*. All this renders plausible the following alternative to the **Explicit Derivation Hypothesis**:

No-Implicature Hypothesis

Judging a sentence of the form *Some X are Y* to be false in a situation where it is obvious that all X are Y does not entail the explicit representation of the reinforced reading *Some, but not all X are Y*

Second, the reconstruction above remains ambiguous as to whether the stronger alternative, *The player scored all of his shots*, needs to be represented. As mentioned above, Katsos and Bishop (2011) merely invoke 'awareness that a more informative statement could have been made'. Strictly speaking, being aware that a more informative statement could have been made does not entail being aware of what this statement could be. More plausibly, though, what Katsos and Bishop (2011) mean is that the stronger *all*-alternative is identified as the content of the more informative statement the speaker could have uttered. In our view, nothing requires that participants who provide such 'pragmatic' responses activate the explicit linguistic representation of this stronger alternative. Everything hinges here on how one constructs step 3, viz. the Gricean Assumption of Quantity, in the reconstruction above. As repeatedly discussed above, Grice's (1989) original formulation focuses on the speaker side—rational speakers do not unnecessarily provide less information than what would be cooperative in the context of conversation. One plausible understanding of step 3, then, is that hearers do not expect speakers to use *Some X are Y* in a situation where it is obvious that all X are Y because their experience tells them that in such situations speakers never do so. In other words, communicative experience should suffice to decide that under-informative *some*-sentences are infelicitous without even explicitly identifying what it would have been cooperative to utter instead. The purpose of this paper, however, is to test the **No-Implicature Hypothesis**, and everything that follows is compatible with both its strong version (under-informative utterances can be rejected without the stronger alternative being represented) and its weaker version (under-informative utterances can be rejected without the stronger alternative being negated).

Third, in addition to supporting the **No-Implicature Hypothesis**, the reconstruction above hints at an interesting contrast between judgements of conversational felicity and judgements of truth. Whereas judging (1a) conversationally inadequate in a situation where the player scored all his shots follows straightforwardly from Gricean assumptions of quantity, judging the same sentence false requires an additional step, which further involves the logical fallacy of Affirming the consequent. We hasten to emphasise that rational reconstructions are just that, and should not be taken at face value as representing psychological processes. That said, there is an intriguing quantitative difference in studies on 'blind' implicatures that use sentence verification tasks: the rates at which under-informative *some*-sentences are judged to be false hover around 30%–40% (Benz & Gotzner, 2014; Geurts & Pouscoulous, 2009; van Tiel & Schaeken, 2017), while the rate at which they are judged to be infelicitous ranges around 80%–98% (Katsos & Bishop, 2011; Papafragou & Musolino, 2003). It makes sense to speculate that this difference owes to the fact that, in this case, truth judgements are parasitic on perceptions of conversational felicity. If the default meaning of *some* is logical, as is largely accepted in the literature, the default judgement of an under-informative sentence is *true*. Accordingly, if judging an under-informative utterance false involves an extra step after having judged it infelicitous, then it should not be surprising that more participants exhibit 'pragmatic' readings of *some* in studies that elicit felicity judgements (i.e. 'infelicitous' response) than in those that rely on truth judgements (i.e. 'false' responses).

1.4. Semantic Error Hypothesis

Benz and Gotzner (2014) take a more radical position on truth judgements of 'blind' implicatures, and claim that those participants who respond 'false' are merely mistaken about the semantic meaning of *some*. The hypothesis, then, is that 'blind' implicatures do not exist:

> **Semantic Error Hypothesis**
>
> Judging a sentence of the form *Some X are Y* to be false in a situation where it is obvious that all X are Y results from mistaking the literal meaning of *some* for *some, but not all* instead of *some, maybe all*

That is, like the **Explicit Derivation Hypothesis**, the **Semantic Error Hypothesis** must assume that the participants who judge under-informative *some*-sentences false associate these sentences with the meaning *some, but not all X are Y*. However, to be genuinely distinguishable from a pragmatic account, the **Semantic Error Hypothesis** must also expect the same misguided participants to associate the quantifier *some* with this same meaning, *some, but not all*, irrespective of the context of use.

1.5. Outlook

As should be clear from the foregoing, even though 'blind' implicatures became a staple of the experimental investigation of scalar implicature, it is not obvious to what kind of representation the critical cases give rise. We have outlined three hypotheses, each involving a different empirical prediction about participants who would judge an under-informative *some*-sentence to be false or infelicitous. The **Explicit Derivation Hypothesis** predicts that such judgements should be associated with the explicit derivation of the corresponding scalar implicature, based on the negation of the stronger alternative. The **No-Implicature Hypothesis**, by contrast, predicts that, in these cases, no explicit representation of the scalar implicature is needed. Finally, the **Semantic Error Hypothesis** predicts that those participants who judge an under-informative *some*-sentence false or infelicitous should associate *some* with the meaning *some, but not all* regardless of the context.

There are several reasons why it is important to test these predictions empirically. To begin with, it is surprising that there are no clear indications as to how one of the most well-known results in experimental pragmatics should be interpreted. To be sure, one may argue that the derivation of the implicature from the literal meaning, understood as a relation between two linguistic representations, should be located at a computational level, without making any commitment as to its actual implementation (see Geurts & Rubio-Fernández, 2015). However, as discussed above, many debates on pragmatic processing presuppose that the behavioural properties of sentence-verification tasks reflect those of the derivation of the scalar implicatures, which clearly endows the **Explicit Derivation Hypothesis** with psychological reality. In this sense, the **Explicit Derivation Hypothesis** also appears to be a cornerstone of grammatical accounts of scalar implicatures—at least if competence properties are assumed to map on empirically observable data. Finally, truth-judgement tasks play a central role in current debates about the possibility of embedded implicatures, the very existence of which would constitute a determining argument in favour of grammatical models (see, e.g. Chemla & Spector, 2011; Geurts & Pouscoulous, 2009; Magri, 2011; Potts et al., 2015; van Tiel et al., 2018).[5]

1.6. Present study

Below we report three sentence-verification studies that set out to adjudicate between these three hypotheses. In line with most the experimental literature, let us dub 'pragmatic response' the judgement that an under-informative *some*-sentence is false or infelicitous; conversely, let us dub 'logical response' the judgement that an under-informative *some*-sentence is true or felicitous. The three studies all consist of two adjacent phases, sentence-verification (Phase 1) and forced choice between two paraphrases of the *some*-sentence (Phase 2). The most central variable measured is the congruence between Phases 1 and 2. Assuming that congruence is not random, all three hypotheses predict that logical responses in Phase 1 should be associated, above chance, with the choice of the *some, maybe all* rather than *some, and not all* explicit paraphrase in Phase 2. As for the pragmatic responses in Phase 1, both the **Explicit Derivation Hypothesis** and the **Semantic Error Hypothesis** predict that they should correlate with the choice of the *some, but not all* rather than *some, maybe all* paraphrase in Phase 2. By contrast, no such prediction follows from the **No-Implicature Hypothesis**. Study 1 measures congruence between truth judgement and explicit paraphrase with a recall task; in Study 2, we explicitly present participants with their initial truth judgement while asking them to

[5] At least, if such embedded implicatures cannot be plausibly attributed to other pragmatic factors.

choose an explicit paraphrase of the *some*-sentence; the final Study 3 is similar but replaces truth judgements with felicity judgements.

Pre-registered methods and materials along with the code files for result analysis and raw data for all the experiments reported below can be found on the Open Science Framework Platform at https://osf.io/c78sd. All the experiments were programmed on PsychoPy2 (Peirce et al., 2019), hosted online on Pavlovia.org and fully administered online via Prolific.

2. Study 1: Implicature memory

The objective of this first experiment was to determine whether once participants make truth-conditional judgements that correspond either to a logical or a pragmatic reading of *some*, they are able to recall to which explicit reading their response corresponded. From now on, let us refer to a sentence of the form *Some X are Y* in a situation where all X are Y as a *target sentence*. If judging such a sentence false corresponds to an explicit representation of the derived implicature, we should observe an above chance association between pragmatic responses to target items and the selection of the sentence that corresponds to this reading in the recall phase. A corollary prediction is that we should observe an above chance association between logical responses to target items and the selection of the sentence that corresponds to this reading in the recall phase. Furthermore, unambiguously true or unambiguously false sentences of the form *Some X are Y* serve as a control to ensure that participants' responses in Phase 2 do not simply correspond to a context-independent lexical meaning of *some*.[6]

2.1. Methods

2.1.1. Study design

Each experimental trial comprised two phases: a sentence-picture verification, of the sort classically used in experimental studies on scalar implicatures (Phase 1), and a recall task (Phase 2). In each experimental trial, Phase 1 was immediately followed by Phase 2.

For Phase 1, the sentence-picture verification task, we used drawn displays in which a player is featured with one or two sets of balls. These displays were inspired by those in Potts et al. (2015) and van Tiel et al. (2018), but we had players of 7 different sports: American football, baseball, basketball, golf, hockey, soccer, and tennis. Each sport was represented by 5 different players. A total of 12 balls featured on every display. That number was chosen to be outside the subitising range (see Potts et al., 2015; van Tiel et al., 2018). Green balls, always on the left, stood for successful shots, while red balls, always on the right, stood for misses. There were 3 possible configurations: a pile of 5 green balls on the left and 7 red ones on the right; a pile of 12 green balls on the left and none on the right; a pile of 12 red balls on the right and none on the left. Below each such display, participants were presented with a sentence built on the following template:

The player scored with all the/no/some balls.

Since we needed a formulation that would be consistent with ways of scoring in all seven sports, we opted for the verb *score* and for placing our quantifiers in a PP-adjunct.[7] We also chose not to use partitive formulations with quantifier followed by *of the*, because these boost pragmatic readings (Degen & Tanenhaus, 2013; Grodner et al., 2010; also Sun & Breheny, 2020). Finally, at the bottom of each Phase 1 screen were displayed two clickable buttons: TRUE, always on the left, and FALSE, always on the right.

All target items had the sentence *The player scored with some balls* underneath a display with 12 green balls. We expected two types of response: a logical one, on which the sentence would be judged to be true, and a pragmatic one, on which it would be judged to be false. Each participant was presented with five such target items. Participants were also presented with another 30 control items with straightforward correct answers. Ten items had the quantifier *all (the)* with either a player and twelve green balls (five unequivocally true cases) or a player with five green and seven red balls (five unequivocally false cases). Ten items had the quantifier *no* with either a player and twelve red balls (five unequivocally true cases) or a player with five green and seven red balls (five unequivocally false cases). Finally, 10 items had the quantifier *some* with either a player with five green and seven red balls (five unequivocally true cases) or a player with 12 red balls (five unequivocally false cases). In the end, each participant was asked to judge 15 correct and 15 incorrect sentence-picture matchings, plus 5 which lent themselves to two mutually incompatible readings. Participants never saw the same player twice. Five sports, those featured in the target items, had an equal proportion of true and false answers (two each). The remaining two sports had either two true and three false answers, or the opposite. We ensured a balanced representation of gender and ethnicity across the depicted sports.

In Phase 2, the recall task, participants were presented with two sentences, together with the question: *Which sentence corresponds more closely to the sentence you saw on the previous screen?* For the items with *all* and *no*, both sentences were identical to the one on the previous screen, except for a mention of the player's sport. In one sentence—e.g. *The hockey player scored with all the balls*—the sport was a correct match for the picture just seen; in the other—e.g. *The soccer player scored with all the balls*—it was not. As for the items with *some*, we provided alternatives that made explicit the meaning of *some*, as either *some, but not all* or *some, maybe all*. In all 15 cases, those alternatives were *The player scored with some, maybe all the balls* and *The player scored with some, but not all the balls*. They were presented in that order in eight of the items, and in the reverse order in the remaining seven. An illustration of experimental trials is provided in Fig. S1.

2.1.2. Procedure and participants

One hundred and one participants were recruited via the Prolific platform; they were pre-screened via Prolific to have English as their first language and not to have taken part in the pre-test to this study (see footnote 6). Participants were instructed that they would have to match sentences with pictures, and that they would be asked some memory questions. In two training items, participants were presented with three geometrical shapes, along with a sentence and asked to click on TRUE or FALSE, to familiarise themselves with the sentence-verification procedure. They moved on to the subsequent screen only after the correct response had been selected. Next, a screen depicting a player with green and red balls introduced participants to the meaning of the ball colours. To move to the next screen and begin the experiment, participants had to click on the balls that the depicted player did not score with, namely the red balls.

2.2. Analytic plan

All statistical analyses were conducted in R. The package lme4 was used for implementing multilevel logistic regressions (Bates et al., 2015), with the package lmerTest for *p*-value estimates (Kuznetsova et al., 2017) and the package emmeans (Lenth et al., 2020) for post-hoc comparisons (with Tukey corrections for multiple comparisons). The effect of fixed factors was assessed in a stepwise fashion, using log-likelihood comparisons between a model with this factor and a model without it but with an otherwise identical structure. During this

[6] A preregistered experiment with an almost identical design served as the pre-test of Study 1 and is available at https://osf.io/5snw7.

[7] The wording in Potts et al. (2015) and van Tiel et al. (2018) – *hit some/all/none of his shots* – was ill-suited, as it only works for a subset of the sports featured. We concede, however, that the improved wording still fell short of being perfectly suited to scoring rules in all seven sports.

stepwise process, the random structure included only by participant random intercepts. Once the best fitting model had thus been identified, it was augmented with the relevant by participant slopes and by item intercepts, unless convergence issues arose. The final models reported below are those with the maximal random structure that allowed convergence.

The first set of models tests the distribution of pragmatic (false) responses to the target items along the trials. The dependent binomial variable is Response-in-Phase-1 (True vs. False), and Trial (viz. trial rank) is the independent variable.

The second set of models investigates the relationship between responses in Phase 1 and Phase 2. For this research question, we included target items, as well as control items with *all* and *no* quantifiers. A binomial Congruence dependent variable was created as follows:

- Control items with *all* and *no*: 1, if the sport selected in Phase 2 matched the sport depicted in the visual stimulus in Phase 1; 0 otherwise.
- Target items: 1, if TRUE was selected in Phase 1 and the form *The player scored with some, maybe all the balls* in Phase 2, or if FALSE was selected in Phase 1 and the form *The player scored with some, but not all the balls* in Phase 2; 0 otherwise.

In other words, in control items with *all* and *no*, congruence measures accurate recall of the sport depicted in Phase 1, while in target items congruence measures the match between Phase 1 response and the reading of *some* to which it is supposed to correspond. Independent variables are Item-type (Control vs. Target) and Trial.

The third set of models focuses on whether Congruence within the target items depends on the response given in Phase 1. The independent variables are Response-in-Phase-1 (True vs. False) and Trial.

Recall, finally, that control items with *some* were unambiguously false or unambiguously true. However, whether, after having been presented with a *some*-control item, in Phase 2, participants selected *The player scored with some, maybe all the balls* or *The player scored with some, but not all the balls* was likely to provide an indication as to how they interpreted *some* in this context. The forced choice of explicit paraphrases of *some*-control items constitutes a further check on the source of the pragmatic or logical readings of target items. If responses to target items in Phase 1 were based on a context-independent meaning of *some*, the same meaning should have emerged when participant were asked to choose between two explicit paraphrases of unambiguously true or unambiguously false *some*-sentences. For this reason, it is also important to compare how these explicit representations of the meaning of *some* in control items compare with the interpretation selected for target items in Phase 1. Accordingly, in the fourth set of models, we created a binomial Logical reading dependent variable, as follows:

- Control items (with *some*): 1, if the sentence selected in Phase 2 is *The player scored with some, maybe all the balls*; 0 if the sentence selected in Phase 2 is *The player scored with some, but not all the balls*
- Target items: 1, if TRUE was selected in Phase 1; 0 otherwise.

The independent variables are Item-type (Some-false, Some-true vs. Target) and Trial.

2.3. Results

2.3.1. Controls

Mean accuracy on control items in Phase 1 ranged from 0.95 to 0.99. We removed all control trials containing errors. Furthermore, seven participants had an error rate above 10% on control items, and were removed from further analyses, leaving a total sample of 94.

2.3.2. Logical vs. pragmatic responses in Phase 1 for target sentences

The mean of pragmatic responses to target items in Phase 1 was 0.76 (sd = 0.43). Fig. 1 displays the average of True responses to target sentences along trials. The addition of Trial to a model with by participant random intercepts improved the model fit ($\chi^2(1) = 5; p = .025$), confirming that the probability of judging a target sentence true rose as the experiment progressed.

2.3.3. Congruence between Phases 1 and 2

As can be seen from Fig. 2, congruence was almost at ceiling for control items with *all* and *no*, but considerably lower for target items. The fit of a model with random by participant intercept was improved by the addition of Item-type (Control vs. Target; $\chi^2(1) = 141.96; p < .001$), but not of Trial and of the Item-type x Trial interaction (both $p > .07$). The model with the random structure augmented with Item-type by participant random slopes (adding by item random intercept did not allow convergence) confirmed that the congruence was lower for Targets than for Controls ($\beta = -1.37; se = 0.39; p = .001$).

Turning to congruence in target items, Fig. 3 confirms that congruence was very high when the response to a target item in Phase 1 was True, and very low when that response was False. Stepwise comparisons of logistic models showed that the addition of Response-in-Phase-1 (True vs. False) improved the model fit ($\chi^2(1) = 271.1; p < .001$), but not Trial or Trial x Response-in-Phase-1 interaction (both $p > .1$). The best fitting model augmented with by item intercepts confirmed that responding True in Phase 1 significantly raises the probability of providing a convergent response in Phase 2 ($\beta = 14.78; se = 2.56; p < .001$). The fitted probability of selecting the logical reading of *some* in Phase 2 after responding True in Phase 1 was 0.99 (se = 0.4×10^{-3}); that of selecting the pragmatic reading of *some* in Phase 2 after responding False in Phase 1 was 0.1×10^{-2} (se = 0.2×10^{-2}).

2.3.4. Readings of *some* in target and control items

Fig. 4 shows that, as in Study 1, the logical readings of *some*-control items in Phase 2 were very much less frequent than the logical responses to target in Phase 1. Stepwise comparisons of logistic models confirmed that the model fit was improved by the addition of Item-type (Some-false vs. Some-true vs. Target; $\chi^2(2) = 953.36; p < .001$) but not Trial or the Item-type x Trial interaction (both $p > .135$). Post-hoc comparisons on the model with by participant random intercepts, which was the maximal model to converge, confirmed that the odds of a logical reading were significantly higher for Target than Control items (Some-false: $\beta = 6.37, se = 0.43; p < .001$; Some-true: $\beta = 6.64, se = 0.45; p < .001$); there was no difference between the two types of control items ($p = .734$).

2.4. Discussion

In this first study, in around 30% of target items, participants judged that a sentence of the form *Some X are Y* is false in a situation where all X are Y. This percentage is very similar to the rate of pragmatic responses that has been reported in other studies on scalar implicatures using truth-judgement tasks (Benz & Gotzner, 2014; Geurts & Pouscoulous, 2009; van Tiel & Schaeken, 2017).

According to the **Explicit Derivation Hypothesis**—which, as argued above, underlies much of the current experimental literature—such pragmatic responses should be taken as an indication that the implicature *Some, but not all X are Y* has been derived. Somewhat more marginally, following the **Semantic Error Hypothesis**, participants who respond pragmatically mistake the literal meaning of *some* for *some, but not all*. Both hypotheses predict that judging that an underinformative *some*-sentence is false should at least favour the activation of a linguistic representation *some, but not all*. Transposed to the context of Study 1, one should expect, then, that this activation should persist from Phase 1 to Phase 2, prompting participants who responded pragmatically in Phase 1 to select the explicit *Some, but not all X are Y*

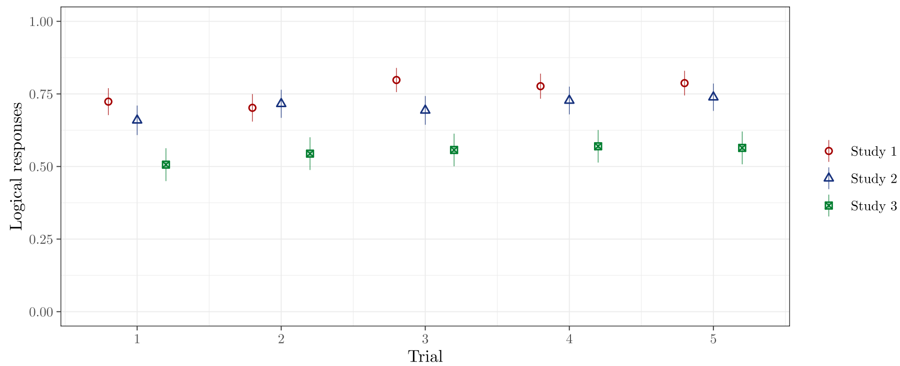

Fig. 1. Mean True (vs. False) response on target items in Phase 1 across trials and studies; vertical bars represent standard errors.

paraphrase in Phase 2. And yet, this is just the opposite of what we observed: independently of their response to target items in Phase 1, the probability for a participant to choose the *Some, but not all X are Y* paraphrase is virtually naught.

It could be objected that what Phase 2 measures is actually not recall but participants' metalinguistic, context-independent intuitions about the meaning of *some*, or, equivalently, that participants mistook the task in Phase 2 for one of providing the logical meaning of the target sentence, and not their interpretation of it. That is, the fact that participants overwhelmingly selected the *Some, maybe all X are Y* paraphrase in Phase 2, independently of their response to targets in Phase 1, would simply indicate that these participants' context-independent interpretation of *some* was *some, maybe all*. However, if this were the default reading of *some*, one would have expected it to be selected also for the *some*-control items, which were unambiguously true or unambiguously false. Yet, for the *some*-control items, in Phase 2, participants overwhelmingly selected the explicit *Some, but not all X are Y* paraphrase, while they did choose the *Some, maybe all X are Y* paraphrase for target items. This makes it rather unlikely that the latter choice reflects a context-independent lexical meaning of *some* that participants would favour.

Another formulation of the **Semantic Error Hypothesis** in the light of the present results could be that participants who provide a pragmatic response to targets in Phase 1 are mistaken about the meaning of *some* but realise their mistake and correct it once they are presented with explicit paraphrases in Phase 2. An argument in favour of this interpretation would be that the rate of logical readings significantly increased along trials. But again, if what participants changed their minds about were the semantics of *some*, we should also have observed a significant increase in the choice of logical readings in Phase 2 for *some*-control trials, which is not what we found. As a matter of fact, participants chose the *Some, but not all X are Y* paraphrase for *some*-control items at a significantly higher rate than that at which they selected the pragmatic response for target items in Phase 1. Therefore, these results also unequivocally indicate that responses to targets in Phase 1 are not based on a meaning of *some* that participants would consistently favour.

On the face of it, a plausible interpretation of the results of Study 1 is that behaving in a way that is compatible with the derivation of a quantity implicature does not involve an explicit representation of this implicature—in line with the **No-Implicature Hypothesis**. Rather, in target trials the participants chose those explicit responses in Phase 2 that most closely matched the situation they had been presented with in Phase 1. Target trials depicted a player who scored with all the balls, and most participants chose the explicit *The player scored with some, maybe all the balls*, which is true of this situation.

As for the *some*-control items that were unambiguously false, the choice of the *The player scored with some, but not all the balls* paraphrase may have been facilitated by the situation pictured in Phase 1, in which the player did not score at all, and therefore made salient that the number of green balls was not even sufficient for *The player scored with some, but not all the balls* to be true. Finally, for the *some*-control items that were unambiguously true, the option *The player scored with some, maybe all the balls* clashes with a pictorial representation that has five green and seven red balls; additionally it is false on the most obvious readings of the modal *maybe*.

All in all, Study 1 indicates that those participants who responded pragmatically did so without having explicitly derived the corresponding implicature. Of course, one could speculate that the representation of the implicature is short-lived and does not persist into the recall Phase 2. Remember, however, that for control items with *all* and *no* the accurate response in Phase 2 required to correctly recollect the sport pictured in Phase 1, and participants were close to ceiling on this measure. Still, it is possible that the representation of the implicature is more fragile than the memory for the sport depicted. Furthermore, in Phase 2 we asked participants about the *meaning* of the sentence displayed on the previous screen. However it is possible that Phase 2 triggers a reinterpretation of the sentence presented relative to the situation depicted rather than tapping the memory of the interpretation it received in Phase 1. There is therefore a risk that a recall task obliterates the implicature that could have potentially been derived in Phase 1. A more stringent test of implicature derivation would be to introduce the explicit paraphrases of Phase 2 while keeping all the information presented in Phase 1 on the screen. This is what we did in Studies 2 and 3.

3. Study 2: Implicature justification

Study 1 showed that participants who judge an under-informative *Some X are Y* sentence as false do not subsequently match their response with the explicitly pragmatic reading of *some*. However, in Study 1, participants had to recall the situation during which they made the truth judgement on a target sentence, and were asked about the meaning of that sentence. It is therefore possible that this recovery process made them focus on the visual stimuli that had been presented and reprocess the sentence anew, which could, in turn, obliterate the linguistic representation of the (putatively) derived implicature. In Study 2, we left both the target sentence and the visual stimuli on the screen when we explicitly asked participants which reading of *some* was a better justification of their response.[8]

3.1. Methods

3.1.1. Study design

The design of this study was identical to that of Study 1, except that Phase 2 was no longer a recall task but an interpretation task. To

[8] A preregistered experiment with an almost identical design served as the pre-test of Study 2, and is available at https://osf.io/mr5cq.

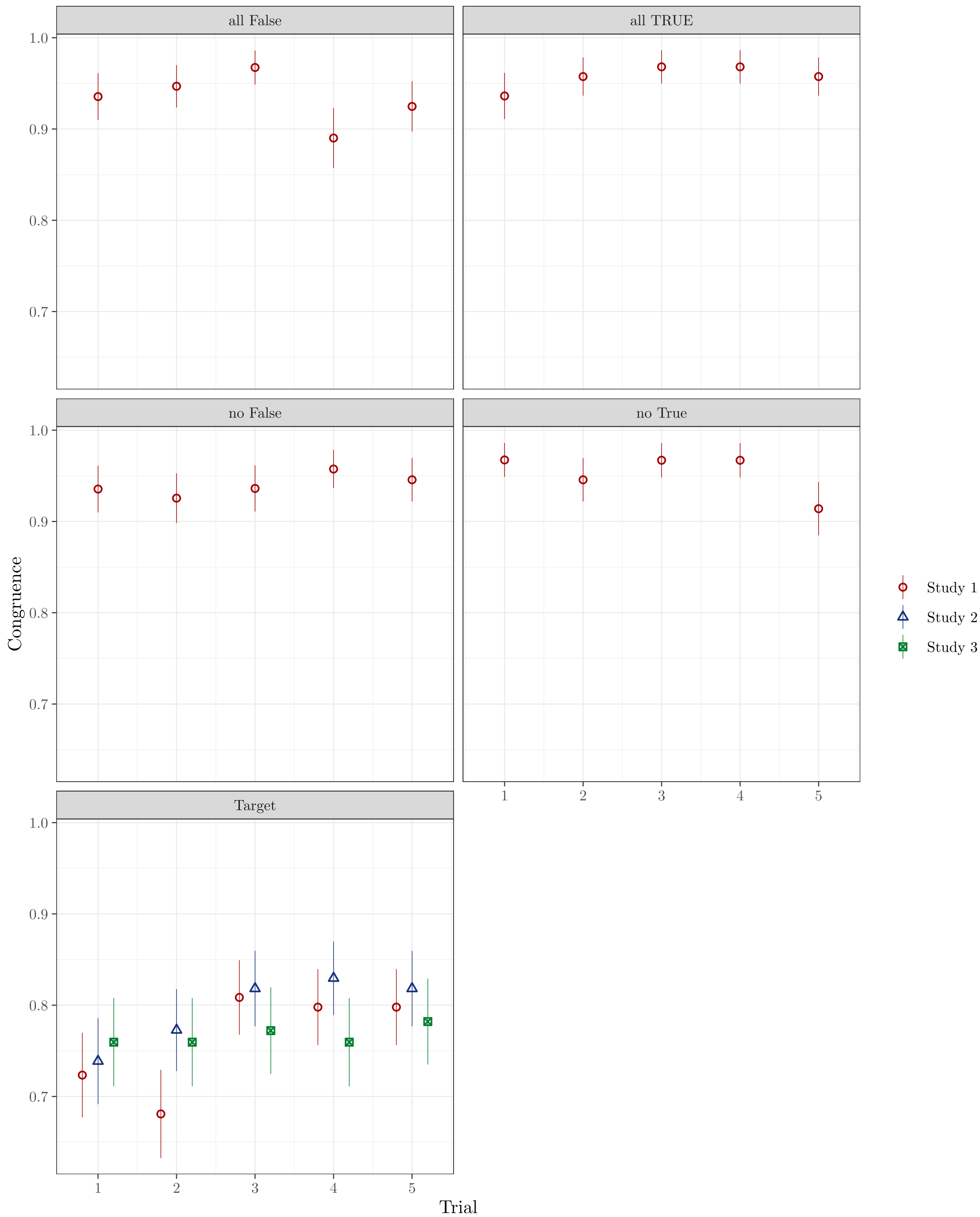

Fig. 2. Mean congruence between Phases 1 and 2 across trials and studies; vertical bars represent standard error. In Study 1, for control items with *all* and *no*, congruence is 1 if, and only if, in Phase 2 the sport in the selected sentence matches that of the picture in Phase 1. For targets, congruence is 1 if, and only if, the reading of *some* in Phase 2 corresponds to the response (logical vs. pragmatic) in Phase 1.

make the interpretation task non-trivial for the control trials with *all* and *no*, we also slightly altered the visual stimuli in Phase 1. We kept the same ball displays as in Study 1 but no longer showed the players. The truth-verification task in Phase 1 was therefore made only relative to the red and green balls displayed on the screen. In Phase 2, the ball display shown in Phase 1 remained on the screen. The Phase 2 screen also displayed a prompt for a forced choice between two sentences. For the *all/no*-controls the prompt was: *The sentence above is about…*. This was followed by two sentences between which the participants had to choose, which were identical to the one on the Phase 1 screen except for a mention of the player's sport. In one sentence—e.g. *The hockey player scored with all the balls*—the sport was a correct match for the ball display; in the other—e.g. *The soccer player scored with all the balls*—it was not. That is, in Phase 2, for *all/no*-controls the foil task consisted in correctly matching a ball display with the correct sport. For the items with *some*—targets and controls— the prompt in Phase 2 was *You decided that the sentence above is TRUE/FALSE because it means that …*, in which the truth-value that appeared in capitals was reproduced from the response just given in Phase 1. In other words, for the *some*-items in Phase 2, participants had their Phase 1 response before their eyes while they chose between two alternatives that made explicit the meaning of *some*: *The player scored with some, maybe all the*

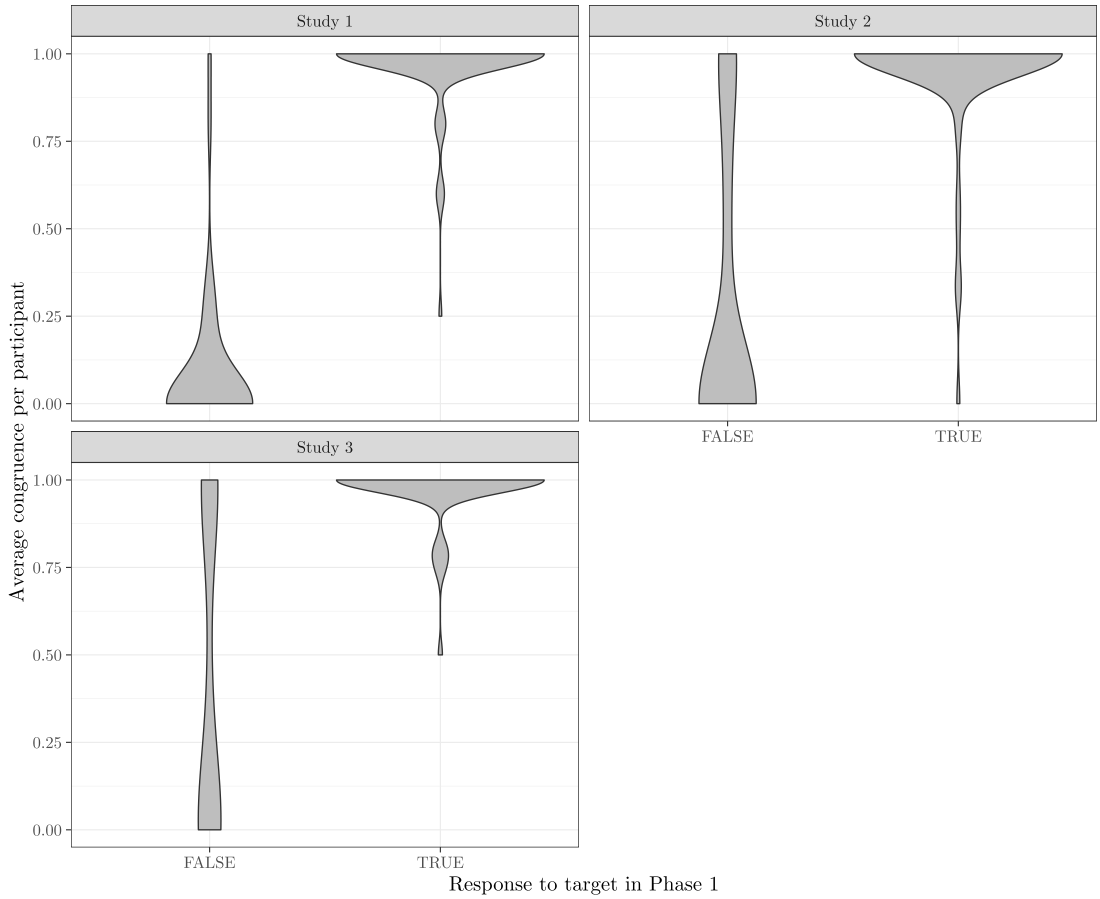

Fig. 3. Distribution of average congruence per participant, as a function of the response to the target in Phase 1, across studies.

balls vs. *The player scored with some, but not all the balls*. An illustration of experimental stimuli is provided in Fig. S2.

3.1.2. Procedure and participants

One hundred and one participants took part in the experiment via the Prolific platform; participants were pre-screened via Prolific to have English as their first language and not to have taken part in Studies 1 or any of the pre-tests (see footnotes 6 and 8). With the provisos just outlined above, the procedure was identical to that of Study 1.

3.2. Analytic plan

The analytic plan is identical to that of Study 1, except that it no longer made sense to compute a congruence score for *all/no*-control items. In Study 1, congruence scores in *all*- and *no*-control items went proxy for the recall of the sentence presented in Phase 1. As just explained, in this study, for the *all/no*-control items in Phase 2 the task was to identify the sport corresponding to the balls displayed on the screen.

3.3. Results

3.3.1. Controls

Mean accuracy on control items in Phase 1 ranged from 0.96 to 0.99. All control trials containing errors were removed. Furthermore, 13 participants had an error rate above 10% on control items, and were removed from further analyses, leaving a total sample of 88.

3.3.2. Logical vs. pragmatic responses in Phase 1 for target sentences

The mean of True responses on target items was 0.71 (sd = 0.45). As can be seen from Fig. 1, the average of True responses to target sentences along trials in Study 2 was very similar to that observed in Study 1. A model with Trial by participant random slopes confirmed that the probability of judging a target sentence true rose as the experiment progressed ($\beta = 1.42; se = 0.47; p = .003$).

3.3.3. Congruence between Phases 1 and 2

As can be seen from Fig. 2, the pattern of congruence for target items in Study 2 resembles that of Study 1. Fig. 3 confirms that congruence was very high when the response to a target item in Phase 1 was True, and very low when the response to a target item in Phase 1 was False. Stepwise comparisons of logistic models showed that the addition of Response-in-Phase-1 (True vs. False) improved the model fit ($\chi^2(1) = 179.62; p < .001$), but not Trial or Trial x Response-in-Phase-1 interaction (both $p > .08$). The best fitting model augmented with by participant random slopes confirms that responding True in Phase 1 significantly raises the odds of providing a convergent response in Phase 2 ($\beta = 18.02; se = 2.51; p < .001$). The fitted probability of selecting the logical reading of *some* in Phase 2 after responding True in Phase 1 was 0.99 (se = 0.1×10^{-3}); that of selecting the pragmatic reading of *some* in Phase 2 after responding False in Phase 1 was 0.2×10^{-3} (se = 0.3×10^{-3}).

3.3.4. Readings of *some* in target and control items

Fig. 4 shows that, as in Study 1, the logical readings of *some*-control items in Phase 2 were very much less frequent than the logical responses to targets in Phase 1. Stepwise comparisons of logistic models confirmed that the model fit was improved by the successive additions

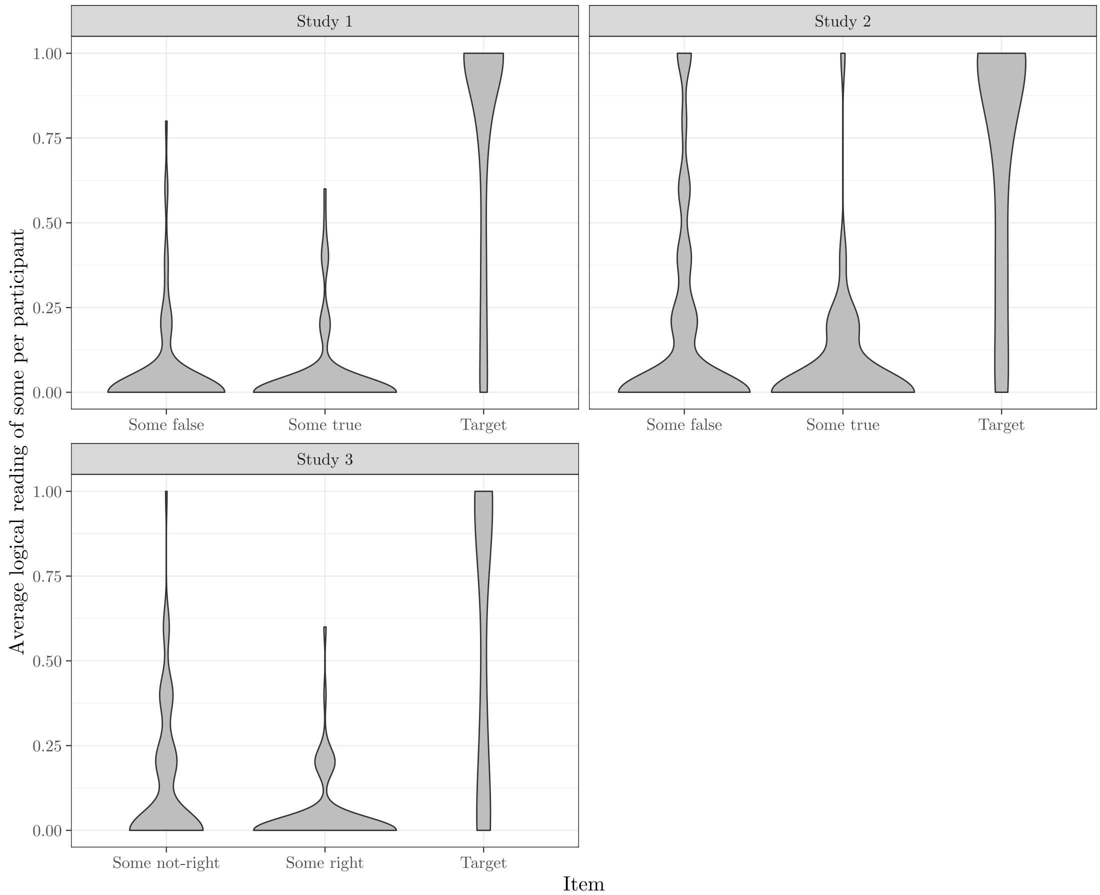

Fig. 4. Distribution of average logical readings per participant across studies. For control items, the logical reading corresponds to the selection of the *The player scored with some, maybe all the balls* sentence in Phase 2; for target items, the logical reading corresponds to the logical response in Phase 1.

of Item-type (Some-false vs. Some-true vs. Target; $\chi^2(2) = 599.01; p < .001$) and Trial ($\chi^2(1) = 5.94; p = .015$), but not by the addition of the Item-type x Trial interaction ($p = .519$). Post-hoc comparisons on the model with Trial by participant random slopes, which was the maximal model to converge, confirmed that the odds of a logical reading were significantly higher for Target than Control items (Some-false: $\beta = 3.57, se = 0.26; p < .001$; Some-true: $\beta = 5.03, se = 0.34; p < .001$), and for Some-false than for Some-true control items ($\beta = 1.46; se = 0.27; p < .001$).

3.4. Discussion

The results of the explicit justification task in Phase 2 of Study 2 fully mirror those of the recall task in Phase 2 of Study 1: independently of whether they provided a pragmatic or a logical response to target items in Phase 1, participants chose the explicit paraphrase that corresponded to the logical reading of *some*. In Phase 2 of Study 2, when selecting the explicit meaning corresponding to their Phase 1 judgement, participants saw on the screen the target sentence, the corresponding ball display, and the truth judgement they had provided about target items. Even then, those who judged a target sentence false almost always selected the *some, maybe all* option in Phase 2. Moreover, like in Study 1, participants did predominantly select the *some, but not all* paraphrase in Phase 2 of *some*-control trials, which confirms that the response pattern in target items cannot be explained as the choice of a context-independent meaning.

Study 2 provides strong evidence in favour of the **No-Implicature Hypothesis**. A residual worry, at this stage, could stem from the fact that, in order to obtain pragmatic readings (and hence provide a window on implicature derivation) Studies 1 and 2 relied on truth judgements. The rate of False responses to target items was remarkably similar across the two studies[9] (and in fact to what has been reported elsewhere), but relatively low, around 30%. This numerical consistency makes it rather unlikely that False responses constitute errors. Nonetheless, it is possible that asking about the truth or falsity of under-informative sentences is not the best way to tap into the derivation of blind scalar implicatures. In this connection, recall that the rates of pragmatic responses to under-informative *some*-sentences are in general higher when they correspond to judgements of infelicity than to judgements of falsity. This difference might be associated with the fact that when asked to provide a pragmatic felicity judgement—rather than a semantic truth judgement—participants do derive the scalar implicature. For this reason, in the next study we replicate the design of Study 2, but replace truth- by felicity-judgements.

4. Study 3: Explicit implicature justification (felicity judgements)

In this final study, we sought to determine whether the absence of congruence between pragmatic judgements of under-informative *some*-sentence and explicit justifications persists when the judgement task focuses on felicity instead of truth.

[9] As well as to their respective pre-tests; see footnotes 6 and 8.

4.1. Methods

4.1.1. Study design

The design of Study 3 is identical to that of Study 2, except that in Phase 1 participants were asked whether the sentence displayed was 'the right thing to say in this situation'. Participants were asked to make this judgement by clicking on YES or NO buttons displayed on the screen. Accordingly, in trials with *some* in Phase 2, participants were presented with the following prompt: *You decided that the sentence above is THE RIGHT/NOT THE RIGHT thing to say in this situation, because it means that …*, in which the 'RIGHT'/'NOT RIGHT' part that appeared in capitals was reproduced from the response just given in Phase 1. In other words, for the *some*-items in Phase 2, participants had their Phase 1 pragmatic response before their eyes while they chose between two alternatives that made explicit the meaning of *some*: *The player scored with some, maybe all the balls* vs. *The player scored with some, but not all the balls*.

4.1.2. Procedure and participants

One hundred and two participants took part in the experiment via the Prolific platform; participants were pre-screened via Prolific to have English as their first language and not to have taken part in Studies 1 and 2, and their pre-tests. With the provisos just described, the procedure was identical to that of Study 2.

4.2. Analytic plan

The analytic plan is identical to that of Study 2, except for two changes in variable coding. First, the Response-in-Phase-1 variable is now coded as Right vs. Not-Right (instead of True vs. False in Studies 1 and 2). Second, the level of the Item-type variable that corresponds to unambiguously true *some* controls was recoded as Some-right (instead of Some-true) and the one corresponding to unambiguously false *some* controls was recoded as Some-not-right (instead of Some-false).

4.3. Results

4.3.1. Controls

Mean accuracy on control items in Phase 1 ranged from 0.88 to 0.96. All control trials containing errors were removed. Furthermore, 23 participants had an error rate above 10% on control items, and were removed from further analyses, leaving a total sample of 79.

4.3.2. Logical vs. pragmatic responses in Phase 1 for target sentences

The mean of Right responses to target items was 0.6 ($sd = 0.5$). As can be seen from Fig. 1, the average of Right, viz. logical, responses to target sentences along trials in Study 3 was somewhat lower than in Studies 1 and 2. Stepwise comparisons of logistic models with by participant random intercepts revealed no effect of Trial on the probability of judging a target sentence Right ($p = .167$).

4.3.3. Congruence between Phases 1 and 2

As can be seen from Fig. 2, the pattern of congruence in target items in Study 3 resembles that of Studies 1 and 2. Focusing on target items, Fig. 3 shows that congruence was very high when the response to a target item in Phase 1 was Right, and low when the response to a target item in Phase 1 was Not-Right. Stepwise comparisons of logistic models showed that the addition of Response-in-Phase-1 (Right vs. Not Right) improved the model fit ($\chi^2(1) = 126.33; p < .001$), but not that of Trial or Trial x Response-in-Phase-1 interaction (both $p > .19$). The best fitting model augmented with by item random intercepts confirms that responding Right in Phase 1 significantly raises the odds of providing a convergent response in Phase 2 ($\beta = 8502; se = 1.63; p < .001$). The fitted probability of selecting the logical reading of *some* in Phase 2 after responding Right in Phase 1 was 0.99 (se = 0.5 $\times$ 10^{-3}); that of selecting the pragmatic reading of *some* in Phase 2 after responding Not-Right in Phase 1 was 0.3 (se = 0.2).

4.3.4. Readings of *some* in target and control items

Fig. 4 shows that the logical readings of *some*-control items in Phase 2 were less frequent than the logical responses to targets in Phase 1. Stepwise comparisons of logistic models confirmed that the model fit was improved by the addition of Item-type (Some-right vs. Some-not-right vs. Target; $\chi^2(2) = 391.93; p < .001$), but not of Trial or of the Item-type x Trial interaction (both $p > .255$). Post-hoc comparisons on the best fitting model with random by participant intercept, which was the maximal model to converge, confirmed that the odds of a logical reading were significantly higher for Target than Control items (Some-not-right: $\beta = 2.94, se = 0.24; p < .001$; Some-right: $\beta = 4.46, se = 0.34; p < .001$), and for Some-not-right than for Some-right control items ($\beta = 1.52; se = 0.33; p < .001$).

4.4. Discussion

The rate of pragmatic responses in Phase 1 was higher in this Study than in Studies 1 and 2, which relied on truth judgements (around 40% instead of 30%), which is in line with the literature. More importantly, we find again that congruence between pragmatic responses and the corresponding explicit justification is very low. The participants who judged that an under-informative *some*-sentence is infelicitous nonetheless largely selected the logical justification for their response. The probability of providing a congruent response in Phase 2 after a pragmatic response in Phase 1 is higher than in previous Studies (around 30% vs. less than 1% in Studies 1 and 2), but still at chance. The design of this study thus offers a particularly stringent test case for the **No-Implicature Hypothesis**. Even though participants were required to make a felicity judgement, and were presented with their judgement together with the target sentence and the corresponding display, those who behaved in a way usually taken to reflect the derivation of a scalar implicature still did not choose the corresponding explicit representation. Note, furthermore, that contrary to Studies 1 and 2 we did not find an effect of Trial on the rate of pragmatic responses, which suggests that being presented with the justification task in Phase 2 did not gradually prevent participants from judging that subsequent under-informative *some*-sentences were infelicitous.

5. General discussion

Blind implicatures have been a staple of experimental pragmatics (and semantics). The almost universal assumption is that when a manifestly under-informative *some*-sentence is judged to be false or infelicitous it is so because the corresponding stronger alternative is negated and the corresponding scalar *some, but not all* implicature is derived. In this series of studies we find clear evidence that such judgements occur without the corresponding implicature being derived. Independently of how participants judge the target sentences, when asked for the meaning of these sentences or for a justification of their judgements, there is a very low probability that they do not opt for the logical *some, maybe all* over the pragmatic *some, but not all* paraphrase. Furthermore, the analysis of explicit justifications of unambiguously true or false *some*-sentences shows that participants strongly favour the pragmatic reading, which confirms that the response pattern we have observed is not due to the fact that they take *some* to mean *some, maybe all* regardless of the context or misunderstand the experimental task as being about the literal meaning of *some*.

In the Introduction, we formulated three distinct hypotheses about blind scalar implicatures:

Explicit Derivation Hypothesis

Judging a sentence of the form *Some X are Y* to be false in a situation where it is obvious that all X are Y entails explicitly representing the reinforced reading *Some, but not all X are Y*

Semantic Error Hypothesis

Judging a sentence of the form *Some X are Y* to be false in a situation where it is obvious that all X are Y results from mistaking the literal meaning of *some* for *some, but not all* instead of *some, maybe all*

No-Implicature Hypothesis

Judging a sentence of the form *Some X are Y* to be false in a situation where it is obvious that all X are Y does not entail the explicit representation of the reinforced reading *Some, but not all X are Y*

The lack of congruence we observed between pragmatic responses and explicit justifications thereof provides strong evidence in favour of the third, **No-Implicature Hypothesis**.

At this point, one might be tempted to explain our results away as simply showing that naive participants cannot be trusted as to their judgements about the meaning they actually assign to linguistic strings—to quantified declarative sentences in the case at hand. By the same token, however, one should also call into question the validity of the same participants' judgements about the truth (or felicity) of these very same strings. There is, indisputably, a metalinguistic or metacognitive component to choosing between two explicit paraphrases of pragmatically ambiguous sentences, but this also applies to the elicitation of intuitions about truth or felicity. And yet, pragmatics, experimental or not—and, for that matter, formal semantics—is predicated on the idea that such judgements provide an accurate window on how we understand and wield language.

A less radical objection would consist in arguing that, at the level of grammatical competence, pragmatic responses to blind implicatures necessarily involve the negation of stronger alternatives and the derivation of the *some, but not all* implicatures, but that this mechanism does not necessarily transpire as a property of performance. A related objection would be that the relationship between the literal meaning of the sentence and the derived implicature should be read as a computational solution to a conversational coordination problem (what the solution should be and why), which bears no commitments at to its algorithmic implementation (how it should unfold; see Geurts & Rubio-Fernández, 2015). An immediate consequence for the field would then be that the relationship between literal *some*-sentences and the implicatures they (putatively) trigger would not be easily transposed to the level of processing. As discussed in the introduction, studies that focus on the processing of blind implicatures presuppose that processing properties reveal something about the link between the literal and the enriched meanings of *some*. In other words, this literature operates on the premise that what it measures is the processing relationship between two linguistic representations (e.g. Bott & Noveck, 2004; Chemla & Singh, 2014b; Marty & Chemla, 2013; van Tiel & Schaeken, 2017, among many others; see Breheny, 2019, for a lucid review).[10] What our results show, however, is that if pragmatic responses to under-informative *some*-sentences do involve the derivation of the *some, but not all* implicatures, this derivation unfolds at a level that never reaches speakers' awareness. While this possibility cannot, of course, be ruled out a priori, a proper defence of it requires a strong empirical motivation to keep framing processing results in terms of implicature derivation. To the best of our knowledge, no such evidence is currently available. Contrastingly, the three studies we report above indicate that responding pragmatically to under-informative *some*-sentences does not seem even to prime participants towards selecting the pragmatically enriched meaning in justifying their decision.

From a theoretical point of view, there must be a strong reason for postulating a link at the competence level that leaves no trace in speakers' awareness. In the generative tradition, the arguments for doing this are usually structural in nature. In the case of implicatures, the strongest argument for assuming that pragmatic readings of *some* are inevitably due to the activation and negation of a stronger alternative comes from so-called embedded implicatures (for explicit statements see, e.g. Chemla & Singh, 2014b; Chemla & Spector, 2011; Magri, 2011; Potts et al., 2015). The sentences at the core of this debate are (dauntingly) complex: the crucial question being whether, for instance, (6a), may give rise to the embedded implicature in (6b), as predicted by conventionalist, grammar-driven approaches (e.g. Chemla & Spector, 2011; Potts et al., 2015) or whether such readings are exceedingly rare and limited to exceptional cases, as predicted by globalist, Gricean approaches (among which Geurts & Pouscoulous, 2009).

(6) a. Exactly one player scored some of his shots
b. ⇝ Exactly one player scored some, but not all of his shots

Such embedded implicatures are at the heart of heated debates which we cannot go into here (see van Tiel et al., 2018, for a detailed investigation and discussion of conflicting results). What is important is that the crucial empirical datum is whether participants actually derive such embedded scalar implicatures (in unmarked cases). And the method used to test this assumption are truth-judgement tasks, exactly of the same kind we used here, but applied to structurally more complex sentences. It is assumed that a participant who judges (6a) false in a situation where one player scored all of his shots and the other players scored none or true in a situation where one player missed some shots, while the others scored all of them is usually taken as evidence that this participant derived (6b). Now, our studies unambiguously show that in simple, non-embedded cases, when participants provide what is supposed to be a 'pragmatic' response, they do not derive the linguistic string corresponding to the scalar implicature. It is therefore doubtful that they would do so in more complex, embedded cases, although this provides a straightforward avenue for future research.

On a more constructive note, the pattern of responses we observed in our studies echoes the justifications provided by pragmatic respondents in Katsos and Bishop (2011) and Papafragou and Musolino (2003), whose rejection of the under-informative *Some X are Y* was based almost entirely on the straightforward fact that, in the situation displayed, all X were Y. Recall that in our studies too participants justified their responses (pragmatic or not) by selecting the logical, enriched meaning *some, maybe all*. It is likely that, in line with Katsos and Bishop's (2011) intuition, what drove participants' judgement of falsity or infelicity was the oddness of using *some* in a situation where all players scored their shots. For this reason, the justification of their decision was probably based on the situation that created this clash. However, as we argued in the introduction, perceiving that the target sentence is odd does not require deriving the implicature; it merely requires knowing that it would be infelicitous to use *some* in such a context. If anything, our explanation of the studies on blind implicatures provides a strong vindication of Gricean rational reconstructions. One clear prediction of Grice (1989) is that we should not perceive as rational a speaker who does not provide sufficient information in a situation where she manifestly can do so (and has no reasons not to). Another way to put this is that, in real life, blind implicatures should not occur, precisely because we would never (barring exceptional reasons which do not apply here) use *Some X are Y* to describe a situation where it is obvious to every party involved that all X are Y.

Now, the fact that pragmatic responses to blind implicatures occur without the negation of the stronger alternative does not entail that in more mundane cases, scalar implicatures are never derived. What our paper does show, however, is that, if it is real, such a derivation is a pragmatic and not a grammatical phenomenon. Recall that on a grammatical view, interpreting *some* necessarily involves weighing

[10] Chemla and Singh (2014b, fnt. 11) do speculate that hearers may somehow store the relationship between *some* and *some, but not all*, and 'use it as a heuristic instead of running through the computations'. Either way, it is presupposed that a participant who judged an explicitly under-informative string 'Some X are Y' sentence false did so because they activated\derived\accessed the string 'Some, but not all X are Y'.

whether the stronger alternative should be negated or not. As abundantly discussed above, our participants did not appear to negate the stronger alternative *all* even if they judged the target sentence to be false. However, our paradigm remains silent as to whether they activated this stronger alternative. As just mentioned, in our opinion, the classic Gricean picture offers everything that is needed to explain the infelicity of under-informative sentences without positing the linguistic representation of any alternatives, negated or not. That said, further (experimental) research is needed to determine whether rejecting an under-informative utterance requires or not representing the content of the more informative alternative.

On a more general note, most research in pragmatics assumes, more or less implicitly, that behavioural and processing properties necessarily map onto (derivational) relationships between linguistic representations—the linguistic meaning and its reconstructed interpretation. For instance, the guiding assumption, proved wrong in this paper, in the literature on scalar implicatures is that processing data bear on the link between the enriched and non-enriched meanings of the target sentences. Recently, the field has somewhat moved away from assuming that pragmatic processing should neatly map on a typology of pragmatic meanings, such as implicature, metaphor, irony, and so on (e.g Deliens et al., 2017, 2018; Kissine, 2016; Wilson et al., 2022). The present paper goes one step further and demonstrates that one should not uncritically hold that pragmatic processing necessarily leads from one linguistic representation to another.

Data availability

Data on OSF. The link is provided in the paper: https://osf.io/c78sd

Acknowledgments

We are very grateful to Valentina Bambini and to two anonymous reviewers for the very constructive and useful feedback on the previsous versions of this paper.

Appendix A. Supplementary figures

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.cognition.2023.105463.

References

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), http://dx.doi.org/10.18637/jss.v067.i01.

Benz, A., & Gotzner, N. (2014). Embedded implicatures revisited: issues with the truth value judgement paradigm. In J. Degen, M. Franke, & N. Goodman (Eds.), *Proceedings of the formal & experimental pragmatics workshop* (pp. 1–6). Tubingen.

Benz, A., & Gotzner, N. (2020). Embedded implicature: what can be left unsaid? *Linguistics and Philosophy*, http://dx.doi.org/10.1007/s10988-020-09310-x.

Bott, L., & Chemla, E. (2016). Shared and distinct mechanisms in deriving linguistic enrichment. *Journal of Memory and Language*, *91*, 117–140. http://dx.doi.org/10.1016/j.jml.2016.04.004.

Bott, L., & Noveck, I. A. (2004). Some utterances are underinformative: The onset and time course of scalar inferences. *Journal of Memory and Language*, *51*(3), 437–457. http://dx.doi.org/10.1016/j.jml.2004.05.006.

Breheny, R. (2019). Scalar implicature. In N. Katsos, & C. Cummins (Eds.), *Oxford handbook of experimental semantics and pragmaticsbook of experimental semantics and pragmatics* (pp. 39–62). Oxford: Oxford University Press.

Carston, R. (2002). *Thoughts and utterances: The pragmatics of explicit communication*. Oxford: Blackwell.

Chemla, E., & Singh, R. (2014a). Remarks on the experimental turn in the study of scalar implicature, Part I. *Language and Linguistics Compass*, *8*(9), 373–386. http://dx.doi.org/10.1111/lnc3.12081.

Chemla, E., & Singh, R. (2014b). Remarks on the experimental turn in the study of scalar implicature, Part II. *Language and Linguistics Compass*, *8*(9), 387–399. http://dx.doi.org/10.1111/lnc3.12080.

Chemla, E., & Spector, B. (2011). Experimental evidence for embedded scalar implicatures. *Journal of Semantics*, *28*(3), 359–400. http://dx.doi.org/10.1093/jos/ffq023.

Cherchia, G. (2004). Scalar implicatures, polarity phenomena and the syntax/pragmatics interface. In A. Belletti (Ed.), *Structures and beyond* (pp. 39–103). Oxford: Oxford University Press.

Degen, J., & Tanenhaus, M. K. (2013). Making inferences: the case of scalar implicature processingtle. In L. Carlson, C. Hölscher, & T. Shipley (Eds.), *Proceedings of the 33rd annual meeting of the cognitive science society* (pp. 3299–3304). Austin, TX: Cognitive Science Society.

Deliens, G., Antoniou, K., Clin, E., & Kissine, M. (2017). Perspective-taking and frugal strategies: Evidence from sarcasm detection. *Journal of Pragmatics*, *119*, 33–45. http://dx.doi.org/10.1016/j.pragma.2017.08.002.

Deliens, G., Antoniou, K., Clin, E., Ostashchenko, E., & Kissine, M. (2018). Context, facial expression and prosody in irony processing. *Journal of Memory and Language*, *99*, 35–48. http://dx.doi.org/10.1016/j.jml.2017.10.001.

Fox, D., & Katzir, R. (2011). On the characterization of alternatives. *Natural Language Semantics*, *19*(1), 87–107. http://dx.doi.org/10.1007/s11050-010-9065-3.

Geurts, B. (2010). *Quantity implicatures*. Cambridge: Cambridge University Press.

Geurts, B., & Pouscoulous, N. (2009). Embedded implicatures?!? *Semantics and Pragmatics*, *2*(4), 1–34. http://dx.doi.org/10.3765/sp.2.4.

Geurts, B., & Rubio-Fernández, P. (2015). Pragmatics and processing. *Ratio*, *28*(4), 446–469. http://dx.doi.org/10.1111/rati.12113.

Grice, P. (1989). *Studies in the way of words*. Cambridge, London: Harvard University Press.

Grodner, D. J., Klein, N. M., Carbary, K. M., & Tanenhaus, M. K. (2010). Some, and possibly all, scalar inferences are not delayed: Evidence for immediate pragmatic enrichment. *Cognition*, *116*(1), 42–55. http://dx.doi.org/10.1016/j.cognition.2010.03.014.

Guasti, M. T., Chierchia, G., Crain, S., Foppolo, F., Gualmini, A., & Meroni, L. (2005). Why children and adults sometimes (but not always) compute implicatures. *Language and Cognitive Processes*, *20*(5), 667–696. http://dx.doi.org/10.1080/01690960444000250.

Jary, M. (2013). Two types of implicature: Material and behavioural. *Mind & Language*, *28*(5), 638–660. http://dx.doi.org/10.1111/mila.12037.

Jary, M. (2016). Rethinking explicit utterance content. *Journal of Pragmatics*, *102*, 24–37. http://dx.doi.org/10.1016/j.pragma.2016.06.003.

Katsos, N. (2009). Evaluating under-informative utterances with context-dependent and context- independent scales: experimental and theoretical implications. In U. Sauerland, & K. Yatsushiro (Eds.), *Experimental semantics and pragmatics* (pp. 51–73). Basingstoke: Palgrave.

Katsos, N., & Bishop, D. V. M. (2011). Pragmatic tolerance: Implications for the acquisition of informativeness and implicature. *Cognition*, *120*(1), 67–81. http://dx.doi.org/10.1016/j.cognition.2011.02.015.

Kissine, M. (2016). Pragmatics as metacognitive control. *Frontiers in Psychology*, *6*(2057), http://dx.doi.org/10.3389/fpsyg.2015.02057.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmertest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *1*(13), http://dx.doi.org/10.18637/jss.v082.i13.

Lenth, R., Singmann, H., Love, J., Buerkner, P., & Herve, M. (2020). Package 'emmeans'. *CRAN Repository*.

Magri, G. (2011). Another argument for embedded scalar implicatures based on oddness in downward entailing environments. *Semantics and Pragmatics*, *4*(6), 1–51. http://dx.doi.org/10.3765/sp.4.6.

Magri, G. (2017). Blindness, short-sightedness, and hirschberg's contextually ordered alternatives: A reply to Schlenker (2012). In S. Pistoia-Reda, & F. Domaneschi (Eds.), *Linguistic and psycholinguistic approaches on implicatures and presuppositions* (pp. 9–54). Cham: Springer International Publishing, http://dx.doi.org/10.1007/978-3-319-50696-8_2.

Marty, P., & Chemla, E. (2013). Scalar implicatures: working memory and a comparison with only. *Frontiers in Psychology*, *4*, http://dx.doi.org/10.3389/fpsyg.2013.00403.

Noveck, I. A. (2001). When children are more logical than adults: experimental investigations of scalar implicature. *Cognition*, *78*(2), 165–188. http://dx.doi.org/10.1016/S0010-0277(00)00114-1.

Noveck, I. A. (2018). *Experimental pragmatics: The making of a cognitive science*. Cambridge: Cambridge University Press.

Papafragou, A., & Musolino, J. (2003). Scalar implicatures: experiments at the semantics–pragmatics interface. *Cognition*, *86*(3), 253–282. http://dx.doi.org/10.1016/S0010-0277(02)00179-8.

Papafragou, A., & Skordos, D. (2016). Scalar implicature. In J. Lidz, W. Snyder, & J. Pater (Eds.), *The Oxford handbook of developmental linguisticsbook of developmental linguistics* (pp. 611–629). Oxford: Oxford University Press.

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo …, H., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. http://dx.doi.org/10.3758/s13428-018-01193-y.

Potts, C., Lassiter, D., Levy, R., & Frank, M. C. (2015). Embedded implicatures as pragmatic inferences under compositional lexical uncertainty. *Journal of Semantics*, *33*(44), 755–802. http://dx.doi.org/10.1093/jos/ffv012.

Schlenker, P. (2012). Maximize presupposition and Gricean reasoning. *Natural Language Semantics*, *20*(4), 391–429. http://dx.doi.org/10.1007/s11050-012-9085-2.

Sperber, D., & Wilson, D. (1995). *Relevance: Communication and cognition* (2nd ed.). Malden: Blackwell Publishing.

Sun, C., & Breheny, R. (2020). Another look at the online processing of scalar inferences: an investigation of conflicting findings from visual-world eye-tracking studies. *Language, Cognition and Neuroscience*, *35*(8), 949–979. http://dx.doi.org/10.1080/23273798.2019.1678759.

van Tiel, B., Franke, M., & Sauerland, U. (2021). Probabilistic pragmatics explains gradience and focality in natural language quantification. *Proceedings of the National Academy of Sciences*, *118*(9), Article e2005453118. http://dx.doi.org/10.1073/pnas.2005453118.

van Tiel, B., Noveck, I., & Kissine, M. (2018). Reasoning with 'some'. *Journal of Semantics*, *35*(4), 757–797. http://dx.doi.org/10.1093/jos/ffy012.

van Tiel, B., & Schaeken, W. (2017). Processing conversational implicatures: Alternatives and counterfactual reasoning. *Cognitive Science*, *41*(S5), 1119–1154. http://dx.doi.org/10.1111/cogs.12362.

Wilson, E., Lawrence, R., & Katsos, N. (2022). The role of perspective-taking in children's quantity implicatures. *Language Learning and Development*, 1–21. http://dx.doi.org/10.1080/15475441.2022.2050236.