Cognition 269 (2026) 106404

Contents lists available at ScienceDirect

Cognition

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

Most pragmatic responses to underinformative *some*-statements are associated with scalar implicatures

Paula Bull-Morales [a,*], Ira Noveck [b], Lewis Bott [a]

[a] The School of Psychology, Cardiff University, 70 Park Place, Cardiff, United Kingdom
[b] Laboratoire de Linguistique Formelle, CNRS and Université Paris Cité, 8 Rue Albert Einstein, 75013, Paris, France

ARTICLE INFO

Keywords:
Psycholinguistics
Pragmatics
Scalar implicatures
Underinformative sentences

ABSTRACT

A common method for investigating scalar implicatures is to use underinformative sentences, such as Some X are Y, when evidence indicates that All X are Y. Underinformative sentences can have a logical (some and possibly all) or pragmatic (some but not all) interpretation. Recently, Kissine and De Brabanter (2023; Cognition) presented experiments that question whether false responses to underinformative sentences indicate explicit derivation of the implicature. Their findings cast doubt on the conclusions of much recent research in experimental pragmatics. Here, we present three experiments that build on their findings using a similar method while incorporating design differences. In a two-phase paradigm, participants evaluated underinformative sentences (Some elephants are mammals) in Phase 1 before selecting a sentence interpretation (logical, pragmatic, or neither) in Phase 2. In all three experiments ($N = 52$; $N = 103$; $N = 100$), participants were congruent (with the explicit derivation hypothesis) significantly more than predictions based on chance, all p's < 0.05, with effects that were more pronounced when considering a subset of participants who were consistent in their Phase 1 interpretations ($N = 22$; $N = 62$; $N = 67$), all p's < 0.05. Overall, the results support the explicit derivation hypothesis, contrary to Kissine and De Brabanter, and are consistent with assumptions in the pragmatics literature. Nonetheless, there are limitations to the paradigm (e.g. the influence of interpretation paraphrasing), and individual differences in consistency with explicit derivation predictions. It is therefore possible that while most participants were deriving enrichments of the sort not all to underinformative sentences, some participants were not.

Grice (1989) famously argued that sentences do not provide enough explicit information for a listener to retrieve the speaker's intended meaning. A listener needs to go beyond the coded meaning of a sentence – through some form of inference or a general application of communicative rules – and enrich it. One case of such enrichments - scalar implicature – has arguably become the most discussed example of an enrichment in the pragmatics literature. In (1) we present an example of a scalar implicature, where (1a) is a sentence uttered by a speaker and (1b) is the commonly accepted enrichment of it:

1. a. Some of the children are in the classroom.

   ⇨ b. Some but not all of the children are in the classroom.

The proposition in (1b) can be derived by assuming the speaker is being truthful and sufficiently informative. As Horn (1972) would put it: if the speaker knew that all of the children were in the classroom, they would have said so; since the speaker did not say *all*, it must not be the case.

For the purposes of the present work, we provide just this thumbnail sketch of scalar implicatures, but it is important to point out that questions about its derivation have led to a rich theoretical literature (e.g. Chevallier et al., 2008; Chierchia et al., 2012; Geurts, 2010; Levinson, 2000; Sauerland, 2005) and that the intensive focus on scalar implicatures has been sustained by experimental investigations that aim to bring to light the linguistic and psychological properties that underlie them (e.g. Bott & Noveck, 2004; Breheny et al., 2006; Chemla & Bott, 2014; Chierchia et al., 2012; Degen & Tanenhaus, 2015; Horn, 1972; Noveck, 2001; Papafragou & Musolino, 2003). Given that scalar implicatures lie at the border of semantics and pragmatics, they are appealing to scholars of language from both of these fields.

Recently, Kissine and De Brabanter (2023; henceforth K&DeB) focused on one classic experimentally-inspired case that is intimately

* Corresponding author.
E-mail addresses: Bull-MoralesP@cardiff.ac.uk (P. Bull-Morales), BottLA@cardiff.ac.uk (L. Bott).

https://doi.org/10.1016/j.cognition.2025.106404
Received 15 July 2024; Received in revised form 20 October 2025; Accepted 9 December 2025
Available online 15 December 2025

associated with investigations of scalar implicature, i.e. when participants are faced with *underinformative* sentences. In such cases, a weak statement is used (and typically as part of a verification task) when its content suggests, or when background information shows, that a stronger version of that sentence is true. Consider a seminal example from Bott and Noveck (2004) in (2), where a proposition is weakened by the presence of an existential quantifier:

(2) Some elephants are mammals.

In this case, the existential quantifier is logically compatible with a stronger (and patently true) one (*All elephants are mammals*) and thus justifies a true response. After all, *all* entails *some*. However, the sentence in (2) could also be considered incompatible with *all* and thus justify a false response. The traditional pragmatic explanation for this is that a false response is taken to indicate that a participant put an upper bound on *some* (to mean *some but not all*), which then distinguishes it from the *all* reading. On this account, a false response thus implies that the participant employed a scalar inference of the sort described in (1). These types of sentences have come to populate the experimental literature precisely because both responses are justified. Put differently, such sentences have the property that if a literal reading of the sentence is applied – something akin to *at least some elephants are mammals* - it justifies (or is compatible with) a *logical* true response, and if an enriched reading of the sentence is applied - something akin to *some but not all elephants are mammals* - it justifies a false response and is considered *pragmatic*. Experiments typically find that participants give both kinds of responses to (2). Provocatively, based on their series of experiments, K&DeB claim that a scalar implicature is not derived when participants respond *false* to underinformative sentences.

K&DeB raise an important question with respect to the interpretation of existing studies on scalar implicature: What is it that people understand when they reply true or false to underinformative sentences? The answer they provide is a useful result for the field but further investigation is warranted. We now take a closer look at their theories and paradigm and present three experiments that build on their findings.

Kissine and De Brabanter (2023).

K&DeB distinguish between the **explicit derivation hypothesis** and the **no implicature hypothesis** as explanations for what it means to respond *false* to underinformative sentences. The explicit derivation hypothesis asserts that judging a sentence of the form *Some X are Y* to be false in a situation where *All X are Y* applies, emerges as a consequence of a *Some, but not all X are Y* interpretation. The no implicature hypothesis asserts that responding false in this situation does not entail an explicit representation of *Some, but not all X are Y*.

They test between these hypotheses by using a two-phase study in which participants first verified statements (Phase 1) and then provided their interpretation of the sentence (Phase 2). In order to explain the paradigm, we focus on K&DeB's Study 1. In Phase 1, participants verified whether sentences were true or false with respect to a picture, which depicts an athlete in front of a set of 12 balls (with green balls to the left indicating scoring success and red balls to the right indicating failure; see Fig. 1). The statements were "The player scored with all balls", "The player scored with some balls" or "The player scored with no balls". Importantly, each sentence in Phase 1 could be classified as true or false as a function of the presentation of the balls. For example, the statement "The player scored with all balls" would be true if all the balls were green, and it would be false if only 7 of the 12 accompanying balls were green or if none of them were. In the second phase, participants were shown two sentences and asked to choose the one that best reflects what they had just read in the previous verification phase ("Which sentence corresponds more closely to the sentence you saw on the previous screen?"). For the control items (concerning *all* and *no*), the task was relatively straightforward: the test sentence from Phase 1 was represented along with a foil that switched the athlete. So, sticking with the baseball example in Fig. 1, Phase 2 would present participants with two options, e.g. *The golfer scored with all the balls* and *The baseball player scored with all the balls*. Participants essentially had to identify the athlete that they had seen in Phase 1.

For the target item (which used a *some* sentence), the task was more subtle: In Phase 1, participants saw an underinformative sentence (e.g. *The player scored with some balls*) when all the balls were presented positively (as green); in Phase 2, participants were presented with two interpretations of the sentence. One reflected an interpretation of the logical reading (*The player scored with some, maybe all the balls*) and the other reflected an interpretation of the pragmatic reading with an explicit negation (*The player scored with some, but not all the balls*).

When participants responded true in the verification phase, they generally selected a logical interpretation. Interestingly, however, when participants responded false in the first phase, they continued to select the logical interpretation at a high rate. K&DeB take this as evidence that participants did not explicitly derive a scalar implicature when they responded false in the underinformative case i.e. evidence against the explicit derivation hypothesis and in favour of the no implicature hypothesis. Their two follow-up experiments largely produced the same outcomes (their Study 2 allowed participants to see the representations of the balls as they carried out Phase 2 and their Phase 1 response, and their Study 3 changed the response options).

The results and conclusion from K&DeB are surprising and important. However, we think the picture is more complex than that presented by K&DeB and that their support for the no-implicature account is premature. In the next section, we discuss some of their data as well as their method in more detail before presenting three experiments that test the same hypotheses as K&DeB while using different materials.

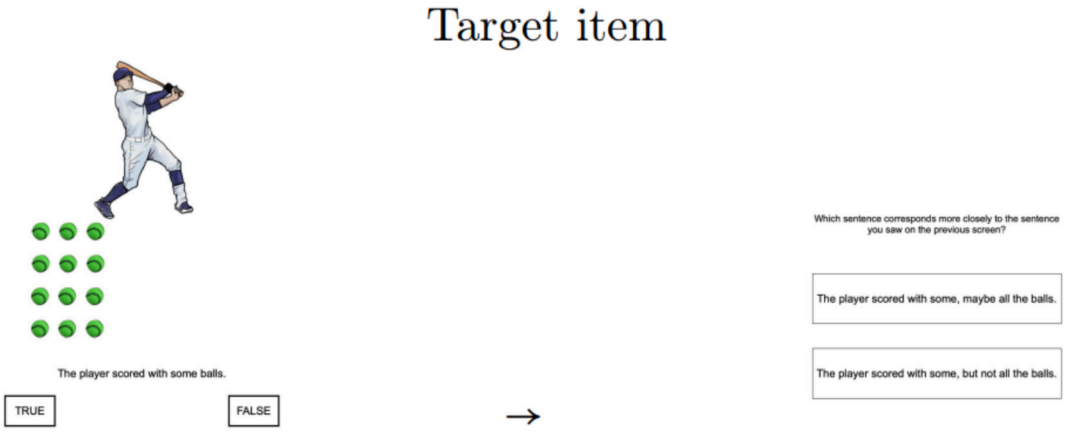

Fig. 1. Target trials during Phase 1 (left) and Phase 2 (right) in K&DeB Study 1.

1. Pragmatic interpretations in K&DeB

K&DeB make strong claims about the absence of association between false responses in Phase 1 and pragmatic interpretations in Phase 2. For example, in the discussion of Study 2, they write, "Those who judged a target sentence false almost always selected the *some maybe all* option [logical interpretation] in Phase 2," and in the discussion of Study 3, "The participants who judged that an underinformative sentence is infelicitous nonetheless largely selected the logical justification for their responses." These claims are reflected in the abstract: "These experiments robustly show that hearers who reject an underinformative sentence do so without explicitly entertaining a *some but not all* implicature." However, the proportion of pragmatic Phase 2 interpretations suggests a more subtle picture than that presented by K&DeB.

Examination of K&DeB's raw data (https://osf.io/c78sd/) shows that in Study 2, 35 % of Phase 2 interpretations were pragmatic after a false Phase 1 response, whereas 4 % of Phase 2 interpretations were pragmatic after a true Phase 1 response. Likewise, in Study 3, where the response options in Phase 1 were felicitous/infelicitous instead of true/false, 48 % of Phase 2 interpretations were pragmatic after a false (infelicitous) Phase 1 response compared to 8 % after a true (felicitous) response. If there were no association between pragmatic interpretations and false responses, the proportion of pragmatic interpretations after false responses would be the same as that after true responses, which they clearly are not.

One reason for the discrepancy between the percentages above and K&DeB's qualitative claims could be that K&DeB focused on the absolute probability of participants selecting the pragmatic interpretation given that they said false in Phase 1, rather than the comparison between the rate of pragmatic responding after a false response and the rate after a true response, which is essentially a control. K&DeB argued that the probability of pragmatic interpretations after a false response was not above chance in their experiments, i.e. participants did not select the pragmatic interpretation significantly more than 50 % of the time after a false response. Since the no implicature hypothesis predicts no relationship between Phase 1 responses and Phase 2 interpretations, i.e. chance responding, they claimed that the data supported the no implicature hypothesis. However, most psycholinguistic experiments do not draw conclusions on absolute rates – absolute rates vary according to the specifics of the sample and the materials - but on comparisons across conditions. In this case, the relevant comparison would be between the rate of the pragmatic responses after false, compared to after true evaluations, which as discussed above is large, e.g. 35 % vs 4 %. When the comparison is considered and not the individual, absolute probability of pragmatic interpretations, the weight of evidence favours the explicit derivation account. After all, if there were no relationship between false responses in Phase 1 and pragmatic interpretations in Phase 2, there should be no difference between the rates of pragmatic interpretations across the different Phase 1 responses.

1.1. Task complexity

The second concern we have is that the task given to participants was overly complex and that this introduced errors and biases into the interpretation judgements of Phase 2.

In Study 1, participants were instructed to choose the interpretation option that "corresponds more closely with the sentence you saw on the previous screen". This required them to understand that "corresponds" refers to the sentence meaning independently of what the sentence refers to (the image). Participants were supposed to disregard the image in their responses and to reproduce their interpretation of the sentence. For example, when selecting an interpretation of "The player scored with some of the balls", they were required to ignore the image of a player scoring with *all* the balls that they had just considered in Phase 1. Deriving a sentence interpretation independent of the referent is a difficult meta-linguistic task that, we argue, was not properly understood by participants, or resulted in errors. Similarly, participants had to recall the sentence and its interpretation from the previous screen. This introduced another potential source of error in Phase 2, as K&DeB note (and which formed the rationale for Studies 2 and 3).

In Studies 2 and 3, participants were no longer asked for correspondence, but to *justify* their previous response: "You decided that the sentence above is TRUE [or FALSE] because it means that…" The problem here is that requiring participants to justify their response post hoc is notoriously unreliable (e.g. Johansson et al., 2008; Nisbett & Wilson, 1977; see Noveck, 2018, Chapter 3; Trouche et al., 2018). For example, in the *blind choice* literature (started by Johansson et al., 2008), participants often make preferences (between two options) and, through a sleight of hand, end up verbally justifying their dispreferred choice seconds later as if it were their preferred one (and without catching on). Current research even takes advantage of this kind of participant obliviousness in tasks similar to K&DeB's. Trouche et al. (2018) presented a quantified negative sentence (e.g. *None of the apples are organic*) and asked participants, in Phase 1, to choose a conclusion that follows (from among those presented in a multiple choice format), as well as to provide an explanation for their choice. In Phase 2, the participant's choice and explanation were presented again in order to be compared to an alternative. As part of a blind choice manipulation, the experimenters would occasionally switch out the participant's own answer (presenting the participant's own explanation as coming from a "previous" participant). Remarkably, only about half of the participants reported noticing when the option they had chosen in Phase 1 was replaced. In Studies 2 and 3 of K&DeB, participants were similarly confronted with two choices in Phase 2 and asked to choose one that justifies their response in Phase 1. We suggest that the Phase 2 answers they provided were largely a consequence of the difficulties of post hoc introspection.

In short, all three studies in K&DeB required participants to engage in complex metalinguistic reasoning. We believe that participants were not necessarily doing the introspective work that was called for. Instead, their responses reflected task difficulty and led to a reliance on the presented image, as we describe in more detail below.

1.2. The role of the image

K&DeB used images depicting real-world scenarios, such as the baseball player (see Fig. 1), and in Phase 1, asked participants to match the sentence to the image. However, in Phase 2, participants were expected to disregard the image and either select the sentence that corresponded most closely to the sentence from the previous trial (Study 1) or justify their response in Phase 1 (Studies 2 and 3). We suggest that participants sometimes matched the interpretation in Phase 2 to the image, either because of difficulties introduced by the complexity of the task or because it was an easier strategy to match to the image than the sentence. If so, participants would be more likely to select the *some, maybe all,* response option, irrespective of how they interpreted the sentence in Phase 1, as described below.

Consider first the target cases. Here, participants were presented with options in Phase 2 that required them to select either the *some but not all* interpretation or the *some, maybe all* interpretation. Participants generally chose the *some, maybe all* option. The image that accompanied these trials was of a player succeeding at *all* the scores (see Fig. 1). Crucially, the *some but not all* interpretation was semantically inconsistent with the image (since the played succeeded at *all* the scores), and so participants who were matching the interpretation to the image would not choose this option. Conversely, the *some, maybe all* interpretation allows the possibility of *all* and is therefore consistent with the image. Participants who matched the response options to the image would therefore select the *some, maybe all* interpretation regardless of whether they derived an explicit implicature in Phase 1. This might explain the preponderance of *some, maybe all* responses irrespective of Phase 1 responses.

For the filler sentences, the majority of answers in Phase 2 for *some*-sentences (see Fig. 4, K&DeB) systematically matched the picture last seen in Phase 1 (or in Studies 2 and 3, the picture that is actually represented as they carry out their choice). For the *some*-true case (which shows the player succeeding at roughly half the scores) participants generally chose the *some but not all* reading in Phase 2. This makes sense given that the (pragmatic) partitive interpretation best captures the partial set picture. For the *some*-false case, in which participants were given a *Player scored with some balls* sentence along with 12 red balls indicating unequivocal failure, the situation is complicated by the fact that both of the available interpretation options are inconsistent with the picture. Nonetheless, given that the pragmatic *some but not all* interpretation at least considers the failures and the logical *some, maybe all* interpretation does not, the former comes closer to matching the picture than the latter. The results from K&DeB's Fig. 4 indeed show that participants have a strong preference for opting for the pragmatic interpretation for *some*-false cases.

Furthermore, for the *no* and *all* filler items, participants were obliged to look at the image to answer the Phase 2 question. Specifically, for the filler items *no* and *all*, the Phase 1 sentence refers to the player, *The player scored with all the balls*, but the Phase 2 interpretation options refer to the specific sports player, *The baseball player scored with all the balls*, vs *The hockey player scored with all the balls*. To answer the question, they had to consult the image rather than the Phase 1 sentence. Given that participants were obliged to use the image to answer some of the filler items correctly, it seems likely that they would continue to base their interpretation on the image for target items.

In summary, we have presented an explanation for K&DeB's findings that differs to the explanation presented by K&DeB. First, we argue that there is evidence from K&DeB's own data for an association between Phase 1 and Phase 2 responses (e.g. 35 % pragmatic interpretations after false; 4 % after true). Second, because of the complexity and structure of the task, we argue that participants were sometimes focusing on the image in Phase 2 rather than the sentence, a strategy which results in a logical interpretation. If we are correct, then the possibility remains that participants are indeed engaged in explicit derivation when they respond false to underinformative sentences, contrary to K&DeB's conclusion.

In the next section we present three experiments that build on K&DeB's study. Our goal in these experiments was not to provide a reply to K&DeB or to test the claims we make above (which we admit are based more on speculation than data), but to investigate the question of how participants understand underinformative sentences without the potential interpretation difficulties associated with K&DeB's original task.

1.3. The current experiments

We conducted a study similar to K&DeB's while using verbal stimuli rather than images. Just as in K&DeB, there were two phases to each experiment. In Phase 1, participants were presented with a target sentence in the style of Bott and Noveck (2004), e.g. *Some elephants are mammals* and were asked whether it was true or false. In Phase 2, participants were asked what the sentence meant.

In addition to changing the stimuli, we made a number of other changes that we thought might facilitate valid Phase 2 interpretations. First, we asked only about sentence meaning, rather than justifications, which would obviate the requirement for participants to engage in post hoc justification about why they had chosen a particular response. Second, we repeated the target sentence in the interpretation options (see Fig. 1). This was so that participants did not have to remember it from the previous trial (c.f. Study 1 of K&DeB) e.g. "Some elephants are mammals MEANS some but not all elephants are mammals". Third, we changed the filler trials so that the sentence interpretation required understanding the quantifier and not just the predicate. This was to encourage deeper processing of the sentences more generally (in K&DeB, accurate performance on many filler trials was possible by only processing the predicate, not the quantifier). Finally, we included an additional interpretation option, "Neither of the above" in case participants did not agree with either of the other interpretations we had provided.

We present three experiments. All three use the basic methodology described above but have slight variations. In Experiment 1, participants were not reminded of their response from Phase 1 (as in Study 1 in K&DeB) but in Experiments 2 and 3 they were (as in Study 2 in K&DeB). In Experiments 1 and 2, we used a different interpretation option to K&DeB. We used "at least some…" instead of "some, maybe all…", as in "Some elephants are mammals MEANS at least some elephants are mammals". In Experiment 3, we reverted to "some and maybe all…", a response option that was similar to K&DeB's. As we show below, these variations produced slightly different results, but the overall pattern remained similar throughout.

Just as in K&DeB, we tested between the explicit derivation hypothesis and the no implicature hypothesis. We also used *congruence* (with the explicit derivation predictions) as the dependent measure. If participants gave a logical Phase 2 interpretation after a true Phase 1 response, or if they gave a pragmatic Phase 2 interpretation after a false Phase 1 response, they received a score of 1 on that trial. Otherwise, they received a zero. All *neither* responses were coded as 0, i.e. incongruent with the explicit derivation hypothesis, regardless of Phase 1 response (at the request of a reviewer). Higher congruence scores correspond to a greater adherence to the explicit derivation hypothesis.

We analysed congruence as a function of Phase 1 response, similar to K&DeB, but we also analysed overall congruence. An overall congruence score that differs significantly from predictions based on chance provides evidence in favour of the explicit derivation hypothesis.

We analysed the data with all participants included and, in an exploratory analysis, with a subset of those who consistently chose either true or false in Phase 1 for underinformative sentences (for similar analyses, see Bott & Noveck, 2004; Dieussaert et al., 2011; Mazzaggio et al., 2021; Noveck & Posada, 2003; Ronderos & Noveck, 2023). We defined consistent participants as those responding 9 out of 10 or 10 out of 10 times true, or false ($p < 0.05$ on a binomial test). This latter analysis was conducted because we wanted a sample of participants who showed evidence of being particularly engaged and motivated in the study (participants who were not motivated or who were confused were unlikely to respond consistently). We also worried that if participants oscillate between true and false in Phase 1, they may be less able to recall their Phase 1 response in Phase 2 (a random response is more difficult to remember than a consistent response), thereby introducing noise into the analysis.

The R code files for analysis and raw data for all three experiments can be found on the Open Science Framework Platform at https://osf.io/qhf2x/.

2. Experiment 1

Experiment 1 tested between the explicit derivation hypothesis and the no implicature hypothesis, as identified by K&DeB. The explicit derivation hypothesis predicts that participants should select a pragmatic interpretation after a false response and a logical interpretation after a true response. In contrast, the no implicature hypothesis predicts no association between Phase 1 response and Phase 2 interpretation.

2.1. Method

Participants. Fifty-two Cardiff University students participated for course credit. Participants were randomly assigned to one of two counter-balancing groups ($N = 29$ Group L, $N = 23$ Group P).

Phase 1. Phase 1 was a sentence verification task in which participants judged whether a sentence was true or false. Sentences were constructed according to the form "[Quantifier][exemplar] are

[category]". There were 10 underinformative trials (*target* trials) and 35 filler trials (see Table 1). Target trials were sentences that used *some* when *all* would have been the more informative quantifier. Filler trials used *some, all*, or *no* as quantifiers, constructed so that there were unambiguous true and false sentences of each. While *no*-sentences are difficult for participants, they were included to ensure that participants processed the quantifier during verification.

Each target sentence used a different exemplar and a different category. Each filler sentence used a different exemplar, but there were overlapping categories.

Target sentences and *all*-true sentences were presented 10 times, and all other sentence types 5 times (Table 1). *All*-true sentences were presented more than other filler sentences because we wished to maximise the number of pragmatic interpretations by making the alternative salient (see Bott & Frisson, 2022; Rees & Bott, 2018).

Sentences were presented in a different random order for each participant.

Phase 2. Phase 2 was designed to elicit interpretations of the sentences seen in Phase 1. Participants saw the sentence from Phase 1 and indicated which interpretation option best matched the meaning of the sentence. Phase 2 was presented immediately after Phase 1.

The sentence from Phase 1 was presented at the top of the screen. Below that, they saw the question, "What did the sentence mean?". They were then presented with three response options. Options 1 and 2 presented different interpretations of the Phase 1 sentence, and Option 3 was always, "Neither of the above.". The structure of the first two options was always, "[Sentence] MEANS [interpretation]." (see Fig. 2).

In target trials, Options 1 and 2 were respectively the logical reading, *At least some X are Y*, or the pragmatic reading, *Some but not all X are Y*. The order of presentation was counterbalanced across subjects. Group L saw the logical interpretation first and the pragmatic second and Group P the reverse. In filler trials, Options 1 and 2 presented different interpretations of the sentence (see Table 2). The order of these was also counterbalanced.

In filler trials, Options 1 and 2 varied depending on the filler trial type (see Table 2).

Procedure. Participants completed a training phase prior to the main phase of the experiment. Training trials comprised only of Phase 1 (sentence verification). Participants saw 12 sentences, two from each filler category (Table 1). Feedback was given following the response, stating whether the selection was "correct" or "incorrect". Target sentences were not used in the training phase.

Participants received instructions before the training and experimental phases. The instructions given prior to the training phase were, "In the first part of the experiment, you will see a sentence and have to indicate whether it is true or false. For example, you might see the sentence "All elephants are mammals". In this case, you would indicate true because all elephants are in fact mammals. After you have made your response, you will get feedback, either "correct" or "incorrect".

In the experimental phase, participants were instructed that they would be making more true/false judgements but without receiving feedback. They were told that after each sentence, they would see a set of options about the sentence meaning. Participants were asked to select the option that best captured how they understood the sentence by pressing the corresponding number on the keyboard (1–3).

Each trial began with a fixation point that was shown for 500 ms, after which a sentence was presented until a true or false selection was made. No time limit was given for a selection to be made. The next trial did not commence until a sentence meaning was selected.

The experiment was programmed in PsyToolKit (Stoet, 2010, 2017).

Analysis Plan. In line with K&DeB, we analysed the congruency between Phase 1 and Phase 2. According to their scoring system, a score of 1 was assigned if a true Phase 1 response was paired with a logical Phase 2 interpretation or a false Phase 1 response was paired with a pragmatic Phase 2 interpretation. Otherwise, a score of 0 was assigned. *Neither* interpretations were treated as incongruent because they are inconsistent with a false-pragmatic and true-logical mapping. The probability of selecting a congruent interpretation was therefore 0.33.

Model-based analyses. We conducted two analyses using the lme4 package in R assuming binary data. The first model included participants and items with random intercepts and was used to test whether the proportion of congruent scores differed from chance (see Equation A).

$$\text{(A)}\quad \text{congruence} \sim 1 + \text{offset}(\text{logodds33}) + (1 \mid \text{subject}) + (1 \mid \text{item})$$

The second model included participants and items with random intercepts and slopes to test whether Phase 1 response was a predictor of congruent scores (see Equation B).

$$\text{(B)}\quad \text{congruence} \sim \text{phase\_1} + (1 + \text{phase\_1} \mid \text{subject}) + (1 + \text{phase\_1} \mid \text{item})$$

The fitted probabilities of selecting a congruent response were derived using the *effects* package in R. Equations A and B are the maximal allowable models given the study design (Barr et al., 2013).

All participants vs *consistent participants.* We conducted analyses on both the complete data set, where participants provided a mixture of true and false responses in Phase 1, and an analysis on participants who only gave consistently true or consistently false responses in Phase 1 (for similar analyses, see Bott & Noveck, 2004; Dieussaert et al., 2011; Mazzaggio et al., 2021; Noveck & Posada, 2003; Ronderos & Noveck, 2023).

Participants were classified as consistent if they responded true or false to 9 or more underinformative sentences in Phase 1 (i.e. $p < 0.05$ on a binomial test). The 90 % consistency threshold allowed for up to 10 % error. Error responses were excluded from analysis and figures.

Data Preprocessing. In Phase 1, participants used the "A" or "L" key to make a response on a QWERTY keyboard, and in Phase 2, the "1","2″ or "3″ key. However, the screen advanced even when participants pressed a key from the non-allowable set. We therefore excluded any trial for which a non-allowed key was pressed, either in Phase 1 or Phase 2. This resulted in the removal of 130 trials (2.8 %).

Table 1
Sentence types and number of presentations.

Sentence type	Example	Number of presentations	True/False response
Target	Some elephants are mammals.	10	True/False
Some False	Some trout are mammals.	5	False
Some True	Some elephants are Indian.	5	True
All True	All elephants are mammals.	10	True
All False	All trout are mammals.	5	False
No False	No elephants are mammals.	5	False
No True	No trout are mammals.	5	True

2.2. Results

Filler Trials. Accuracy to filler trials in Phase 1 was high overall ($M = 0.91$, $SD = 0.29$) but *no*-true sentences were more difficult than *some*- or *all*-sentences, as is commonly reported in the literature (e.g. Bott & Frisson, 2022). There was large variation in the distribution of interpretation judgements in Phase 2 to filler trials (see Table 2). For some sentences, one interpretation was overwhelmingly selected by participants (e.g. *no*-true), but for others, choices were distributed evenly across the three response options (e.g. *some*-false).

Target Trials. Participants responded true in 58 % of trials. Following a false Phase 1 response, more pragmatic interpretations were selected than any other (74 %), but following a true Phase 1 response,

1. Some elephants are mammals MEANS at least some elephants are mammals.
2. Some elephants are mammals MEANS some but not all elephants are mammals.
3. Neither of the above.

Fig. 2. Interpretation options for the "Some elephants are mammals" target sentence.

Table 2
Filler sentences. Accuracy (Phase 1) and interpretations (Phase 2) in Experiment 1 (E1), Experiment 2 (E2) and Experiment 3 (E3). Option 3 interpretation was always "Neither of the above.". Interpretation proportions exclude interpretations made following an incorrect Phase 1 true/false judgement.

Sentence Type	Phase 1 Accuracy			Interpretation Option 1	Proportion Option 1			Interpretation Option 2	Proportion Option 2			Proportion Option 3		
	E1	E2	E3		E1	E2	E3		E1	E2	E3	E1	E2	E3
All False	0.95	0.99	0.98	Every X is a Y.	0.68	0.59	0.43	No X is a Y.	0.26	0.34	0.44	0.06	0.07	0.13
Some False	0.94	0.97	0.97	At least some X are Y/ Some and maybe all X are Y.	0.35	0.30	0.12	Some but not all X are Y.	0.41	0.37	0.42	0.24	0.33	0.46
No False	0.92	0.93	0.92	Every X is a Y.	0.06	0.09	0.11	There are no X that are Y.	0.74	0.67	0.51	0.20	0.24	0.38
All True	0.92	0.91	0.93	Every X is a Y.	0.88	0.87	0.96	No X is a Y.	0.03	0.03	0.03	0.10	0.10	0.02
Some True	0.94	0.92	0.90	At least some X are Y/ Some and maybe all X are Y.	0.31	0.31	0.18	Some but not all X are Y.	0.68	0.69	0.81	0.01	0.00	0.01
No True	0.78	0.74	0.77	Every X is a Y.	0.05	0.03	0.03	There are no X that are Y.	0.94	0.95	0.95	0.01	0.02	0.03

more logical interpretations were made than any other (50 %) (Fig. 3).

All Participants. Overall congruence ($M = 0.60$, $SD = 0.49$) was higher than chance (0.33), $\beta = 1.38$, $SE = 0.33$, $z = 4.20$; $p < 0.001$ (see Equation A), indicating support for the explicit derivation hypothesis. Congruence in target items was significantly higher when the Phase 1 response was false than when it was true, $M = 0.74$ ($SD = 0.44$) vs $M = 0.50$ ($SD = 0.50$) (Fig. 4), $\beta = 1.43$, $SE = 0.51$, $z = 2.82$; $p = 0.005$ (Equation B), illustrating that congruency differed across true and false Phase 1 responses. However, unlike K&DeB, congruency was higher after a false Phase 1 than a true Phase 1 response. The fitted probability of selecting a pragmatic interpretation following a false Phase 1 response was 0.93, 95 % CI [0.66, 0.99], whereas the probability of making a logical interpretation after a true Phase 1 response was 0.44, 95 % CI [0.25, 0.65].

Consistent Participants. Twenty-two participants (42 %) responded with 90 % consistency in Phase 1. Of those, 6 responded false and 16 responded true. Following false Phase 1 responses, only pragmatic interpretations were made, and following true Phase 1 responses, a logical interpretation was the most frequent (Fig. 3).

Congruence for consistent participants was significantly higher than chance, $\beta = 4.63$, $SE = 2.18$, $z = 2.12$; $p = 0.034$, consistent with the explicit derivation hypothesis. Congruence following a false Phase 1 response was at ceiling ($M = 1$, $SD = 0$) and considerably higher than congruence following a true Phase 1 response ($M = 0.65$, $SD = 0.48$) (Fig. 4). No difference was observed in congruency following a false Phase 1 response and a true Phase 1 response, $\beta = 10.52$, $SE = 159.01$, $z = 0.07$; $p = 0.947$. The fitted probability of selecting a pragmatic interpretation following a false Phase 1 response was 1 (all 6 participants responded false 100 % of the time), and the fitted probability of selecting a logical interpretation following a true Phase 1 response was

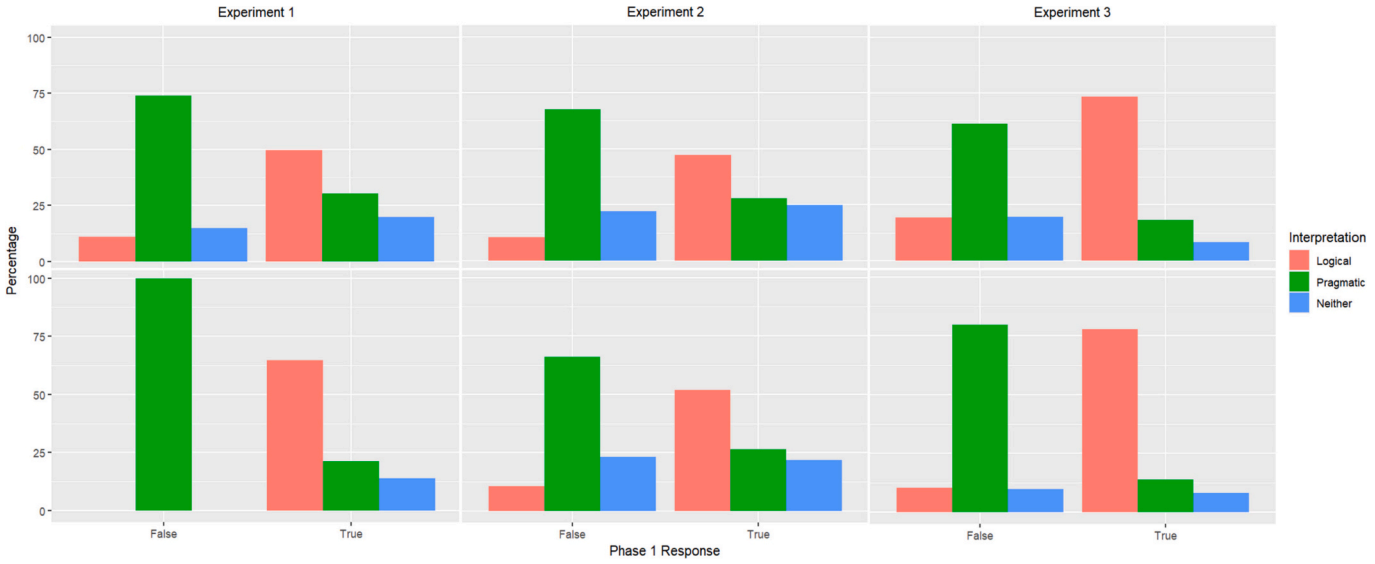

Fig. 3. The percentage of interpretation choices of target trials in Phase 2 as a function of Phase 1 response across experiments. The first row displays interpretation choices of all participants and the second displays interpretation choices from consistent participants only. False responses account for 42 %, 41 % and 34 % of the Phase 1 trials across the three experiments respectively (all participants). Note that the raw percentages reported are nested under participants and items. In all panels, pragmatic interpretations are the most frequent interpretation after false Phase 1 responses, and logical interpretations are the most frequent after true Phase 1 responses.

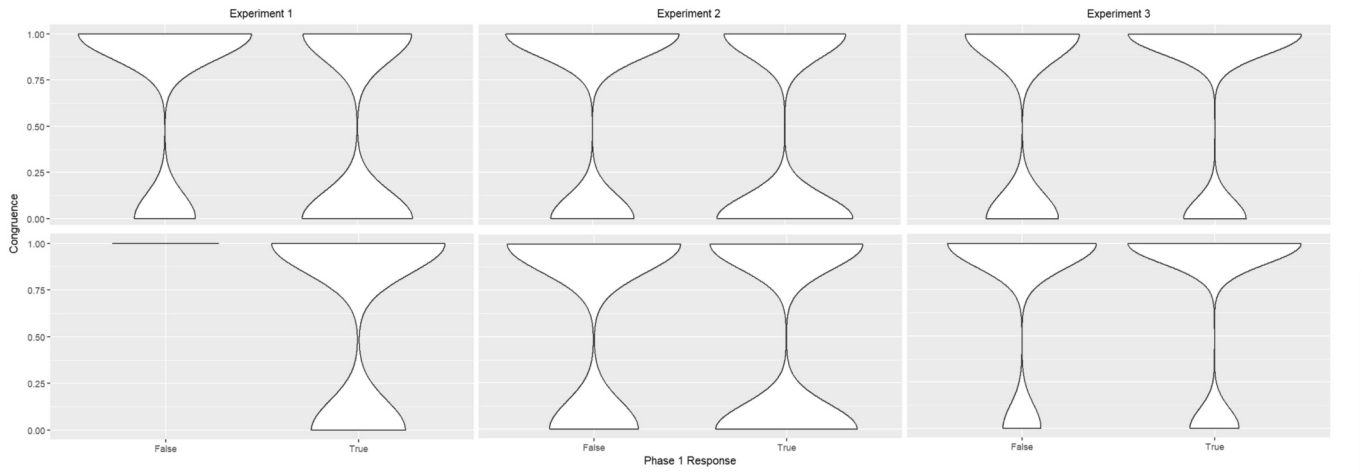

Fig. 4. Congruency scores for Phase 1 responses across three experiments. The first row displays congruence of all participants and the second displays congruence from consistent participants only. Note that congruency for consistent participants in Experiment 1 following a false response is at ceiling.

0.82, 95 % CI [0, 0.97].

2.3. Discussion

When participants responded true in Phase 1, they chose logical interpretations more often than any other, and when they responded false, they chose pragmatic interpretations more than any other. Overall levels of congruency were significantly higher than chance, and when true and false Phase 1 responses were considered separately, congruency for false scores was significantly above chance (but not congruency for true responses). When only consistent participants were considered, those who responded false in Phase 1 always responded with a pragmatic interpretation, consistent with the explicit derivation hypothesis, but those who responded true varied between logical and pragmatic interpretations. Given the significant overall congruency and high rate of pragmatic interpretations in false responders, the evidence supports the explicit derivation hypothesis.

There were nonetheless departures from the most straightforward predictions of the explicit derivation hypothesis. Specifically, there were higher proportions of *neither* and pragmatic interpretations for true responses than might be expected from random errors, and indeed congruence was significantly higher for false responses than true responses. The proportion of consistent participants (42 %) was also low. These data suggest variation in the consistency of mapping from Phase 1 responses to Phase 2 interpretations.

In Experiment 2 we sought to reduce this variability by reminding participants of their Phase 1 response when making their Phase 2 interpretations. By reminding participants we hoped to remove the variability caused by participants forgetting their Phase 1 response.

3. Experiment 2

Participants completed the same sentence verification task as Experiment 1 but were reminded of their Phase 1 response in Phase 2. As in Experiment 1, the explicit derivation hypothesis predicts that logical interpretations should follow true responses and pragmatic interpretations should follow false responses. The no implicature hypothesis predicts no association between Phase 1 response and Phase 2 interpretation.

3.1. Method

Participants. One hundred and three Cardiff University students participated for course credit. Participants were randomly assigned to one of two counter-balancing groups ($N = 50$ Group L, $N = 53$ Group P).

Design, Procedure and Analysis. All components of the design were the same as for Experiment 1, except that participants were reminded of their Phase 1 response in Phase 2. Following their Phase 1 response, participants were presented with the sentence, "You selected the sentence to be TRUE [FALSE]. What do you think the sentence means?" along with the three interpretation options (see Fig. 2).

Data Preprocessing. Fifty-four trials (0.6 %) were removed due to incorrect keypress (see Experiment 1 Data Preprocessing).

3.2. Results

Filler Trials. Accuracy to filler trials in Phase 1 was high overall ($M = 0.91$, $SD = 0.29$). Consistent with Experiment 1, *no*-true sentences were less accurate than *some*- or *all*-sentences and variation in Phase 2 interpretations was observed (see Table 2). Some sentence interpretations were selected more frequently than any other (e.g. *no*-true), whilst others were divided across the three interpretation options (e.g. *some*-false).

Target Trials. Participants responded true in 59 % of trials in Phase 1. Following a false Phase 1 response, a pragmatic interpretation was the most common interpretation (68 %). Following a true Phase 1 response, a logical interpretation was most common (47 %) (Fig. 3).

All Participants. Overall congruence ($M = 0.56$, $SD = 0.50$) was above chance (0.33), $\beta = 1.05$, $SE = 0.23$, $z = 4.68$; $p < 0.001$ (Equation A), supporting the explicit derivation hypothesis. Congruency was higher following a false Phase 1 response ($M = 0.68$, $SD = 0.47$) than a true Phase 1 response ($M = 0.47$, $SD = 0.50$) (Fig. 4), $\beta = 0.91$, $SE = 0.26$, $z = 3.42$; $p < 0.001$ (Equation B). The fitted probability of selecting a pragmatic interpretation following a false Phase 1 response was 0.81, 95 % CI [0.63, 0.91]. In contrast, the probability of choosing a logical interpretation after a true Phase 1 response was 0.40, 95 % CI [0.27, 0.55].

Consistent Participants. Sixty-two participants (60 %) responded with 90 % or higher consistency in Phase 1. Of which, 19 responded false and 43 responded true. Following a false Phase 1 response, a pragmatic interpretation was selected most often. Following a true Phase 1 response, a logical interpretation was selected most often (see Fig. 3).

Consistent with the explicit derivation hypothesis, overall congruence ($M = 0.56$, $SD = 0.50$) was above chance, $\beta = 1.26$, $SE = 0.39$, $z = 3.20$; $p = 0.002$ (see Equation A). There was no significant difference between congruency after a false Phase 1 response ($M = 0.66$, $SD =$

0.48) compared to a true Phase 1 response (M = 0.52, SD = 0.50) (Fig. 4), β = 0.51, SE = 0.45, z = 1.13; p = 0.260. The fitted probability of selecting a pragmatic interpretation following a false Phase 1 response was 0.78, 95 % CI [0.46, 0.94] and the probability of making a logical interpretation after a true Phase 1 response was 0.57, 95 % CI [0.34, 0.77].

3.3. Discussion

The results of Experiment 2 were similar to those of Experiment 1. When participants judged sentences as true in Phase 1, they chose logical interpretations most often, and when they judged sentences as false, they chose pragmatic interpretations most often. Overall, congruence was significantly above chance. Congruence for false responses alone was significantly above chance when all participants were included and when only consistent participants were included. Congruence for true responses was not significantly above chance when all participants were included but was significantly above chance when consistent participants were included (unlike Experiment 1). Data therefore support the explicit derivation hypothesis.

Reminding participants of their Phase 1 response elevated the number of consistent participants from 42 % in Experiment 1 to 60 % in Experiment 2, suggesting that overall variability was lower when participants were reminded of their Phase 1 response in Phase 2. However, there remained a high proportion of *neither* and pragmatic responses after a true response in Phase 1 (Fig. 3). One potential explanation is that participants were relatively unfamiliar with the "At least some…" construction (as opposed to "Some but not all…") and so selected the other interpretations if they were uncertain. We therefore conducted another experiment in which we used "Some and maybe all…" instead of "At least some…".

4. Experiment 3

The design of Experiment 3 was identical to that of Experiment 2 except that "Some and maybe all…" was used as the logical interpretation instead of "At least some…". The predictions for the explicit derivation and the no implicature account remain the same.

4.1. Method

Participants. One hundred participants were recruited using Prolific and were randomly assigned to one of two counter-balancing groups (N = 42 Group L, N = 58 Group P).

Design, Procedure and Analysis. All components of the design were identical to Experiment 2 except that in Phase 2, the wording of the logical interpretation was changed from "At least some" to "Some and maybe all". This was the case for the target sentences and for relevant filler items (see Table 2).

Data Preprocessing. Thirty-seven trials (0.4 %) were removed due to incorrect keypress (see Experiment 1 Data Preprocessing).

4.2. Results

Filler Trials. Accuracy to filler trials in Phase 1 was lower than in Experiments 1 and 2 but high overall (M = 0.86, SD = 0.35). *No*-true sentences were the most difficult for participants to judge correctly in Phase 1 (see Table 2). Phase 2 interpretations continued to be varied with some sentences having clear interpretation selections (e.g. *all*-true) and others being evenly distributed across interpretation options (e.g. *some*-false).

Target Trials. Participants responded true on 66 % of Phase 1 trials. As in previous experiments, pragmatic interpretations were selected most often after a false response (61 %) and logical interpretations were selected most often after a true response (74 %) (Fig. 3).

All Participants. Overall congruency (M = 0.69, SD = 0.46) was above chance, β = 2.05, SE = 7.26 × 10^{-4}, z = 2830; p = 0.034) (see Equation A). Unlike Experiments 1 and 2, congruence was higher following a true Phase 1 response (M = 0.74, SD = 0.44) than following a false Phase 1 response (M = 0.61, SD = 0.49)(Fig. 4), β = −1.32, SE = 0.47, z = −2.81; p = 0.005 (Equation B). The fitted probability of selecting a pragmatic interpretation following a false Phase 1 response was 0.20, 95 % CI [0.05, 0.55]. In contrast, the probability of choosing a logical interpretation after a true Phase 1 response was 0.78, 95 % CI [0.68, 0.86].

Consistent Participants. Sixty-seven participants (67 %) responded with 90 % consistency in Phase 1. Of which, 18 responded false and 49 responded true. Following a false Phase 1 response, a pragmatic interpretation was selected most often, and following a true Phase 1 response, a logical interpretation was the most selected (Fig. 3).

Congruency for consistent participants was high (M = 0.78, SD = 0.41) and above chance, β = 3.50, SE = 0.51, z = 6.82; p < 0.001. Congruence following false (M = 0.80, SD = 0.40) and true Phase 1 responses (M = 0.78, SD = 0.42) were high (Fig. 4), and there was no difference between them, β = 0.49, SE = 0.52, z = 0.83; p = 0.407. The fitted probability of selecting a pragmatic interpretation following a false Phase 1 response was 0.97, 95 % CI [0.82, 1] and the fitted probability of choosing a logical interpretation after a true Phase 1 response was 0.93, 95 % CI [0.82, 0.98].

4.3. Discussion

Participants again most often selected logical interpretations after true responses and pragmatic interpretations after false responses. Overall congruency scores were consequently significantly higher than chance. This indicates support for the explicit derivation account. However, when Phase 1 responses were considered separately, the result was less clear. When all participants were analysed, the probability of logical interpretations following true responses were significantly greater from chance, but the probability of pragmatic interpretations following false responses was not. When only consistent participants were analysed, logical interpretations significantly followed true responses and pragmatic interpretations significantly followed false responses (see Fig. 3), and to a high degree (the probability of selecting a pragmatic interpretation following a false Phase 1 response was 0.97). Support for the explicit derivation hypothesis was therefore strong for the consistent participants but not for those participants who were uncertain.

Despite the similarities in results between Experiment 3 and Experiments 1 and 2, the use of the "some and maybe all" paraphrase instead of "at least some" likely influenced interpretation choices, at least for those participants who were uncertain. The confidence intervals around pragmatic interpretations indicate that there was a lower probability of selecting pragmatic interpretations after false in Experiment 3 compared to Experiment 2, but a higher probability of selecting logical interpretations after true responses in Experiment 3 compared to Experiment 2. In short, the "some and maybe all" paraphrase encouraged a higher rate of logical interpretations after responding true (25 % higher compared to Experiment 2; see Fig. 3), and a lower rate of pragmatic interpretations after responding false (8 % lower). We discuss reasons for this difference in the GD.

5. General discussion

This study tested whether participants who believed underinformative sentences were false also believed those sentences carried an implicature. Our results suggest that they do. In three experiments, we found that participants who responded false in Phase 1 provided pragmatic interpretations more often than any other interpretation in Phase 2, whereas those who responded true in Phase 1 provided logical interpretations more often than any other in Phase 2 (Fig. 3). This was shown by congruency scores at significantly above chance levels in all

experiments, both when all participants were analysed and when only a subset of the most consistent Phase 1 participants were analysed. We therefore find support for the explicit derivation hypothesis, consistent with underlying assumptions in the pragmatics literature (e.g. Bott et al., 2012; Bott & Frisson, 2022; Bott & Noveck, 2004; Marty et al., 2013; Marty & Chemla, 2013; Noveck, 2001; Noveck & Posada, 2003; Ronderos & Noveck, 2023; Schaeken et al., 2018; Van Tiel & Schaeken, 2017).

"Neither of the above". While the overall pattern of responses supports the explicit derivation account, there was a surprisingly high rate of "neither of the above" interpretations for true (20 % Experiment 1; 25 % Experiment 2; 8 % Experiment 3) and false (15 % Experiment 1; 22 % Experiment 2; 20 % Experiment 3) Phase 1 responses. One explanation for this is that subtle differences between the interpretation wording and participant intuitions could have led participants to reject the logical and pragmatic interpretations and favour "neither of the above" instead. Another potential reason was that participants might have believed the sentences were ambiguous between the two specific meanings. They therefore felt they could not choose the logical or pragmatic interpretation, the only remaining option being "neither of the above".

To what extent is the high rate of neither interpretations inconsistent with either the explicit derivation hypothesis or the no implicature hypothesis? We feel that "neither of the above" interpretations do not distinguish between the two theories. Both theories could argue that a mismatch between the internal representations of the sentence meaning and the experimenter-defined paraphrase give rise to "neither of the above" interpretations after either true or false Phase 1 responses. Nonetheless, the high rate of "neither of the above" indicates a limitation of the paradigm, and ultimately, it is possible that participants who respond "neither of the above" are not deriving an implicature when they respond false, i.e. consistent with the no implicature account suggested by K&DeB.

5.1. Departures from the explicit derivation account

When we considered all participants (consistent and inconsistent), we observed a high rate of pragmatic interpretations after true responses (30 % Experiment 1; 28 % Experiment 2; 18 % Experiment 3), and a high rate of logical interpretations after false responses (11 % Experiment 1; 10 % Experiment 2; 19 % Experiment 3). This pattern is inconsistent with the explicit derivation hypothesis.

One potential explanation of this is that participants who gave logical interpretations after saying false in Phase 1 (11 %, 10 %, 19 % respectively) were simply not deriving the implicature, exactly as the no implicature account predicts. This is possible, but it doesn't explain why participants responded with pragmatic interpretations in Phase 2 after saying true in Phase 1 (30 %, 28 %, 18 %). The no implicature account predicts logical interpretations regardless of Phase 1 response and so offers no explanation for why participants would ever systematically give pragmatic interpretations.

We think the likely cause of reverse mapping responses (aside from participant error) is that participants can change their interpretation between Phase 1 and Phase 2. The problem is that sentence truth and interpretation are measured at different time points. Participants could be assessing the sentence anew in Phase 2, possibly coming up with a different interpretation, or they could change their interpretation after further reflection and reading the interpretation options. This represents a core limitation of our study and that of K&DeB (although note that the significant congruence effects – a measure of the match between Phase 1 and Phase 2 - means that the number of occasions when re-interpretation occurred was comparatively small).

Comparison with K&DeB. The results that we obtained are different when compared to those of K&DeB. In particular, we found that participants responded with a pragmatic interpretation after false responses (74 % Experiment 1; 68 % Experiment 2; 61 % Experiment 3) more frequently than a logical one (11 % Experiment 1; 10 % Experiment 2; 19 % Experiment 3), but in K&DeB, participants responded with a higher rate of logical interpretations after false responses than pragmatic interpretations. Here we consider the differences between our paradigms and what might have led to the difference in results.

One difference between our studies was that we used world knowledge verification statements whereas K&DeB used compatibility with a salient image. As we stated in the Introduction, we think participants in K&DeB matched Phase 2 sentences to the image (which depicted *all* and so was not consistent with the *not all* interpretation) rather than to the linguistic structure of the sentence. This was made more likely by the use of *all* and *no* filler items that had to be answered by looking at the image in Phase 2 rather than the sentence in Phase 1. Using world knowledge and different filler items in our study eliminated reliance on the salient image in Phase 2 and so lowered the rate of logical interpretations.

A second difference between our two studies was in the Phase 2 wording. Participants in our task were asked what the sentence in the previous trial meant and then saw interpretation options of the form "[X] MEANS [Y]" e.g., "Some elephants are mammals MEANS at least some elephants are mammals." Participants in K&DeB were asked which of two options justified their response in the previous trial (Studies 2 and 3) or which of two options corresponds more closely to the sentence they saw in the previous trial (Study 1), together with the logical and pragmatic interpretations. We chose the format for Phase 2 because we felt it was easier for participants to understand what was required of them than trying to remember the previous sentence (K&DeB Study 1), to recall their introspection, or to justify their response (K&DeB Study 2 and 3). If we are correct, this could have removed response biases (e.g. responding to the image) and allowed more valid on the spot interpretation judgements at two timepoints.

In contrast to this argument, a reviewer suggested that not explicitly linking the Phase 2 interpretation with Phase 1 would mean participants derived a metalinguistic meaning of the sentence, rather than linking the interpretation to the Phase 1 response. We feel that there are metalinguistic judgements required in both paradigms, however. In our study, by identifying the meaning of a sentence, and in K&DeB's, by requiring justification of the Phase 1 response or judging the closeness of two sentences. We also do not see any evidence that participants were using their Phase 1 response less in our task than K&DeB's (we observed significant congruence between Phase 1 and Phase 2, which is only possible if participants use the Phase 1 response to inform Phase 2).

In summary, we argue that the use of world knowledge and question format was primarily responsible for the difference in the overall rate of pragmatic interpretations across studies. However, we accept that there are a number of differences between the two paradigms, and we have not systematically tested which factor, or which combination of factors, explains the difference in results. What we hope is that the data presented here provides a sufficient demonstration that the use of a different paradigm can lead to different results with respect to the observed proportion of pragmatic interpretations accompanying false responses. The generality of K&DeB's claims – that participants almost always select logical interpretations irrespective of whether they believe the verification sentence to be true or false – should be interpreted in light of our results.

5.2. Paraphrasing of interpretations

As discussed in Experiment 3, there are divergent congruence patterns between Experiment 2, where we used "at least some" and Experiment 3, where we used "some and maybe all". Although we found significant congruence in both experiments, there were 27 % more logical interpretations after true responses in Experiment 3 than Experiment 2, and 8 % fewer pragmatic interpretations after false responses in Experiment 3 than Experiment 2. For true responses, participants seemed to have favoured "some and maybe all" in Experiment 3 instead of "neither of the above," as evidenced by a drop in the rate of

"neither of the above" in Experiment 3 after true responses.

This pattern could be explained in two ways. The first is that participants didn't understand "at least some" but nonetheless believed the sentence meant *some and maybe all*. Thus, when faced with the target sentence that they were uncertain about, they selected "neither of the above" rather than the incomprehensible expression. This effect was then eliminated in Experiment 3, where a more comprehensible paraphrase was used. The second is that "some and maybe all" appeared more consistent with the correct version of the sentence (e.g. *All elephants are mammals*; or "The player scored with all of the balls" in K&DeB) than " at least some" (perhaps because the former contains the word "all") and so if participants were uncertain how to answer Phase 2, they may have chosen "some and maybe all" because it matched the correct version.

The explanations above relate to the biases that arise when people make interpretation judgements. In contexts where the sentence is difficult to understand and the interpretations subtle, participants will be swayed by relatively superficial factors such as the degree to which they are familiar with the expression. Furthermore, participants will vary in how they internally represent the meanings of sentences and so it is difficult to obtain a single paraphrase that all participants will agree with. We don't see this as problematic for evaluating our hypotheses however. What is important is not the overall frequency with which a particular interpretation is selected (which will indeed vary with the paraphrasing) but the change in interpretation choices that arise after the manipulation of interest. In our case, this change is represented by the overall congruency score, which significantly favours the explicit derivation hypothesis in all three experiments.

Filler sentences. Rates of correct answers to filler sentences in Phase 1 were high and in line with previous findings (e.g. Tomlinson Jr et al., 2013). However, there was large variation in the Phase 2 interpretations of filler sentences. There seem to be several factors that contribute to this variability. First, interpretations seem to reflect world knowledge as well as linguistic interpretations. For example, for *all*-false sentences, e.g. *All trout are mammals*, the *every* interpretations (*Every trout is a mammal*) accounted for 68 % of interpretations in Experiment 1 and the *no* interpretations (*No trout is a mammal*) for 26 %. While the *every* interpretation is linguistically the most similar to the target sentence, the *no* interpretation reflects world knowledge. Likewise, the high rate of "neither" responses for *no*-false sentences (20 % Experiment 1; 24 % Experiment 2; 38 % Experiment 3), e.g. *No elephants are mammals*, could reflect a bias against selecting the *no* interpretation (*There are no elephants that are mammals*) on the grounds that the *no* statement is not consistent with world knowledge. Thus, when participants didn't know what the correct interpretation was, they were influenced by their world knowledge.

The second factor contributing to the variability is that for some sentence types, there was no obviously correct intended interpretation. For example, for *some*-true, e.g. *Some elephants are Indian*, the intended meaning of the sentence could be the logical or the pragmatic interpretation, and both interpretations resulted in consistency between the sentence and world knowledge. In these cases, there was no reason to expect participants to overwhelmingly select one interpretation.

It is tempting to make a direct comparison with filler sentence interpretations in K&DeB and note that interpretations in K&DeB were highly consistent across participants. However, filler sentences in K&DeB were quite different from those in our task. For *all*- and *no*-sentences in K&DeB, interpretation options involved a sentence with the same or different subject to the target sentence (and the image), e.g. for an image with hockey players, one interpretation was *The hockey player scored all the goals*, and the other was *The soccer player scored all the goals*. Participants simply needed to match the player in the sentence to the player in the image to identify the intended interpretation, and they had no need to consider the quantifier, unlike our filler items.

Implications for theory. A number of theoretical consequences stem from our results. The most important relates to claims from previous studies on implicature processing. As noted by K&DeB, the majority of studies on scalar implicatures assume that the processing properties of underinformative sentences reveal something about the link between the literal and strengthened meanings of *some* (Bott et al., 2012; Bott & Frisson, 2022; Bott & Noveck, 2004; Marty et al., 2013; Marty & Chemla, 2013; Noveck, 2001; Noveck & Posada, 2003; Ronderos & Noveck, 2023; Schaeken et al., 2018; Van Tiel & Schaeken, 2017). In the absence of *some but not all* interpretations to underinformative sentences, K&DeB question the validity of this literature.

Our findings go some way to restoring the plausibility of the claims behind these studies. In our task, participants responded with pragmatic interpretations after false responses significantly more often than logical interpretations, which suggests that when participants responded false to underinformative sentences, they were deriving implicatures. However, we note that there were considerable individual differences in interpretations, e.g., 20 % of responses after false in Experiment 3 were "neither of the above," rather than pragmatic, contrary to what the explicit derivation hypothesis predicts. Conclusions regarding scalar implicature processing and underinformative sentences should therefore be caveated with the recognition that a minority of participants may not be deriving implicatures or are not confident in their judgements.

6. Conclusion

An important question in research on scalar implicatures is why participants respond false to underinformative sentences. According to standard assumptions in the experimental pragmatics literature, people respond false because they derive a *some but not all* interpretation. When K&DeB tested this, however, they took their results to mean that there is no association between false responding and the derivation of the implicature and so argued against this assumption. Our results paint a more complex picture. Across three experiments, we find significant support for the explicit derivation hypothesis. In particular, participants who consistently responded false to underinformative sentences were significantly more likely to choose a *some but not all* interpretation of an underinformative sentence than any other interpretation. However, there are clear limitations to the two-phase paradigm (e.g. the influence of interpretation paraphrasing), and there are individual differences with respect to how people interpret underinformative sentences (e.g. the high rate of "neither of the above" interpretations). It may therefore be the case that while the majority of participants derive scalar implicatures to underinformative sentences, as our study shows, there is a minority who do not.

CRediT authorship contribution statement

Paula Bull-Morales: Writing – review & editing, Writing – original draft, Resources, Methodology, Investigation, Formal analysis, Conceptualization. **Ira Noveck:** Writing – review & editing, Conceptualization. **Lewis Bott:** Writing – review & editing, Writing – original draft, Supervision, Resources, Project administration, Methodology, Data curation, Conceptualization.

Acknowledgements

This work was funded by a Cardiff University PhD Scholarship.

Data availability

Data and code available at: https://osf.io/qhf2x/.

References

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255–278.

Bott, L., Bailey, T. M., & Grodner, D. (2012). Distinguishing speed from accuracy in scalar implicatures. *Journal of Memory and Language, 66*(1), 123–142.
Bott, L., & Frisson, S. (2022). Salient alternatives facilitate implicatures. *PLoS One, 17*(3), Article e0265781.
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.
Breheny, R., Katsos, N., & Williams, J. (2006). Are generalised scalar implicatures generated by default? An on-line investigation into the role of context in generating pragmatic inferences. *Cognition, 100*(3), 434–463.
Chemla, E., & Bott, L. (2014). Processing inferences at the semantics/pragmatics frontier: Disjunctions and free choice. *Cognition, 130*(3), 380–396.
Chevallier, C., Noveck, I. A., Nazir, T., Bott, L., Lanzetti, V., & Sperber, D. (2008). Making disjunctions exclusive. *The Quarterly Journal of Experimental Psychology, 61*(11), 1741–1760.
Chierchia, G., Fox, D., & Spector, B. (2012). The grammatical view of scalar implicatures and the relationship between semantics and pragmatics. *Semant.:Int. Handbook Nat. Language Meaning, 3*, 2297–2332.
Degen, J., & Tanenhaus, M. K. (2015). Processing scalar implicature: A constraint-based approach. *Cognitive Science, 39*(4), 667–710.
Dieussaert, K., Verkerk, S., Gillard, E., & Schaeken, W. (2011). Some effort for some: Further evidence that scalar implicatures are effortful. *The Quarterly Journal of Experimental Psychology, 64*(12), 2352–2367.
Geurts, B. (2010). *Quantity implicatures*. Cambridge University Press.
Grice, H. P. (1989). *Studies in the way of words*. Harvard University Press.
Horn, L. R. (1972). *On the semantic properties of logical operators in English*. Los Angeles: University of California.
Johansson, P., Hall, L., & Sikström, S. (2008). From change blindness to choice blindness. *Psychologia, 51*(2), 142–155.
Kissine, M., & De Brabanter, P. (2023). Pragmatic responses to under-informative some-statements are not scalar implicatures. *Cognition, 237*, Article 105463.
Levinson, S. C. (2000). *Presumptive meanings: The theory of generalized conversational implicature*. MIT press.
Marty, P., Chemla, E., & Spector, B. (2013). Interpreting numerals and scalar items under memory load. *Lingua, 133*, 152–163.
Marty, P. P., & Chemla, E. (2013). Scalar implicatures: Working memory and a comparison with only. *Frontiers in Psychology, 4*, 51438.
Mazzaggio, G., Panizza, D., & Surian, L. (2021). On the interpretation of scalar implicatures in first and second language. *Journal of Pragmatics, 171*, 62–75.
Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84*(3), 231.
Noveck, I. (2018). *Experimental pragmatics: The making of a cognitive science*. Cambridge University Press.
Noveck, I. A. (2001). When children are more logical than adults: Experimental investigations of scalar implicature. *Cognition, 78*(2), 165–188.
Noveck, I. A., & Posada, A. (2003). Characterizing the time course of an implicature: An evoked potentials study. *Brain and Language, 85*(2), 203–210.
Papafragou, A., & Musolino, J. (2003). Scalar implicatures: Experiments at the semantics–pragmatics interface. *Cognition, 86*(3), 253–282.
Rees, A., & Bott, L. (2018). The role of alternative salience in the derivation of scalar implicatures. *Cognition, 176*, 1–14.
Ronderos, C. R., & Noveck, I. (2023). Slowdowns in scalar implicature processing: Isolating the intention-reading costs in the Bott & Noveck task. *Cognition, 238*, Article 105480.
Sauerland, U. (2005). DP is not a scope island. *Linguistic Inquiry*, 303–314.
Schaeken, W., Van Haeren, M., & Bambini, V. (2018). The understanding of scalar implicatures in children with autism spectrum disorder: Dichotomized responses to violations of informativeness. *Frontiers in Psychology, 9*, Article 348157.
Stoet, G. (2010). PsyToolkit: A software package for programming psychological experiments using Linux. *Behavior Research Methods, 42*, 1096–1104.
Stoet, G. (2017). PsyToolkit: A novel web-based method for running online questionnaires and reaction-time experiments. *Teaching of Psychology, 44*(1), 24–31.
Tomlinson, J. M., Jr., Bailey, T. M., & Bott, L. (2013). Possibly all of that and then some: Scalar implicatures are understood in two steps. *Journal of Memory and Language, 69*(1), 18–35.
Trouche, E., Johansson, P., Hall, L., & Mercier, H. (2018). Vigilant conservatism in evaluating communicated information. *PLoS One, 13*(1), Article e0188825.
Van Tiel, B., & Schaeken, W. (2017). Processing conversational implicatures: Alternatives and counterfactual reasoning. *Cognitive Science, 41*, 1119–1154.