Distinctions between primary and secondary scalar implicatures To appear in *Journal of Memory and Language*

Anouk Dieuleveut[a 1] ; Emmanuel Chemla[b] ; Benjamin Spector[c]

[a] : Institut Jean Nicod (ENS — EHESS — CNRS)
Département d'études cognitives, École normale supérieure - PSL Research University
University of Maryland

[b] : Laboratoire de Sciences Cognitives et Psycholinguistique (ENS, EHESS, CNRS)
Département d'Études Cognitives, École Normale Supérieure, PSL Research University

[c] : Institut Jean Nicod (ENS — EHESS — CNRS)
Département d'études cognitives, École normale supérieure - PSL Research University

Keywords: Pragmatic reasoning, quantity implicature, primary scalar implicatures, secondary scalar implicatures, epistemic step, scalar diversity.

Abstract

An utterance of *Some of the students are home* usually triggers the inference that it is not the case that the speaker believes that all students are home *(Primary Scalar Implicature)*. It may also license a stronger inference: that the speaker believes that not all students are home (*Secondary Scalar Implicature*). Using an experimental paradigm which allows to distinguish between these three distinct readings as such (*literal reading, primary SI, secondary SI*), we show that the secondary SI can be accessed even in contexts where the speaker is *not* presented as being well-informed, a result which goes against classical neo-Gricean pragmatic approaches to Scalar Implicature, but is compatible with both the 'grammatical' approach to Scalar Implicatures and some more recent game-theoretic pragmatic models in which speakers and listeners engage in sophisticated higher-order reasoning about each other. Second, we use this paradigm to compare standard scalar items such as *some* and expressions whose interpretation has been argued to involve SIs, but controversially: *almost*, numerals and plural morphology. For *some* and *almost*, we find that speakers do access three distinct readings, but for numerals and plural morphology, only the literal reading and the secondary implicature could be detected, and no primary implicature, which suggests that the pragmatic and semantic mechanisms at play are different for both types of items.

General Audience Summary

A sentence such as *Some of the students are home* is generally interpreted as meaning that *some, but not all, of the students are home,* even if the sentence is logically true when all students are home. The 'but not all' inference associated to *some* is standardly viewed as a *Scalar Implicature* (SI) and is taken to arise because a sentence containing *some* competes with another (more informative) sentence the speaker could have chosen instead, *All of the students are home*. In this paper, we focus on a more fine-grained distinction made by current theories of SIs, that goes beyond the distinction between the literal reading and the reading with a scalar implicature: between the *primary scalar implicature* - inference that *it is not the case that the speaker believes that all of the students are home*, and the *secondary scalar implicature* - stronger inference that *the speaker believes that not all of the students are home*. Using a novel experimental paradigm which allows to distinguish between these three readings as such (literal reading, reading with just a primary implicature, reading with a secondary implicature), we test the inferences triggered by a variety of linguistic expressions which have been argued to trigger scalar implicatures, but controversially: *some*, *almost*, numerals and plural morphology. We find that (i) for all of these items, the reading with a secondary implicature can be accessed even in contexts where the speaker is not presented as being well-informed, which goes against standard pragmatic theories where secondary implicatures are assumed to arise only when it can be assumed that the speaker is well-informed (*Competence Assumption*), and (ii) that scalar items differ with respect to the accessibility of these readings: if in the case of *some* and *almost*, we detected the three readings, in the case of numerals and plural, only two readings could be detected: no interpretation with just a primary implicature was observed. We discuss this finding in light of current debates about the differences between diverse alleged scalar items.

---

[1] Permanent address:
Linguistics Department, University of Maryland, College Park
anouk.dieuleveut@gmail.com

Introduction and Background

A sentence such as (1) can be understood to convey either (i) that, according to the speaker, there are students who are home, without any implication as to whether all students are home (*literal meaning*), (ii) that the speaker believes that some students are home, but it is not the case that she believes that all students are home (reading with a *primary scalar implicature*) or (iii) that the speaker believes that some students are home, and she believes that it is not the case that all students are home (reading with a *secondary scalar implicature*)[2].

(1) Some of the students are home.

There are currently two main theoretical approaches to the derivation of primary and secondary implicatures. According to *pragmatic, neo-Gricean approaches*, the *primary scalar implicature* (PSI) results from a reasoning about the speaker's communicative goal: a speaker who used (1) rather than (1') signals that she does take the alternative sentence (1') to be true, since otherwise she would have used (1') which is strictly more informative. This is so because <*some, all*> form a so-called *Horn scale* (Horn, 1972), i.e. a sentence with *some* is automatically compared with its counterpart with *all*.

(1') All of the students are home.

This conclusion is compatible with two states of affairs: the speaker could believe (1') to be false, or merely be uncertain as to its truth-value. The *secondary scalar implicature* (SSI) then derives from a strengthening of this primary implicature: if the hearer believes that the speaker is *opinionated* about the more informative sentence (1'), i.e. can be taken to know its truth-value, the hearer reasons that the speaker must take (1') to be false (Sauerland's 2004 *opinionated speaker assumption*, van Rooij, & Schulz's 2004 *competence assumption*, Zimmermann's and Geurts' *Authority Assumption*).

Table 1. Three interpretations for sentences containing scalar items.

Literal meaning (LIT)	B_S (1)	The speaker believes that (1) is true.
With primary SI (PSI)	$B_S (1) \wedge \neg B_S (1')$	The speaker believes that (1) is true and it is not the case that the speaker believes that (1') is true.
With secondary SI (SSI)	$B_S (1) \wedge B_S (\neg 1')$	The speaker believes that (1) is true and the speaker believes that (1') is not true.

According to the alternative *grammatical approach* to SIs (Chierchia, 2004, 2006; Chierchia, Fox, & Spector, 2009, 2012; Fox, 2007, a.o.),[3] the SSI derives from the presence of an optional *exhaustivity operator* (which we will notate *O*) in the logical form of the sentence, whose meaning is similar to *only* (*Only some of the students are home* means that some but not all of the students are home). The derivation of the SSI is thus independent from that of the PSI. Under this view, the primary reading is also an option, and involves either a pragmatic strengthening similar to the Neo-Gricean mechanism, applied to the literal reading (first parse below - applied to the second parse, a similar reasoning would generally not produce any enrichment, Fox, 2007, 2014), or from a third possible logical form (Meyer, 2013, 2014). Crucially, the strengthening takes place after the disambiguation choice the addressee has to make between the two parses: even if both the primary and the secondary readings are predicted, the latter is not an enrichment of the former.

- Parse without the operator O: (1) = *Some, and possibly all, students are home*. The addressee further applies pragmatic mechanisms (maxim of quality) and obtains what corresponds to the literal meaning under the Gricean approach.
- Parse with the operator O: O(1) = *Only some of the students are home*., i.e., *Some, but not all, students are home*. This parse leads to Bs (O(1)), which corresponds to the SSI: $B_S ((1) \wedge \neg (1'))$.

According to both approaches, the relevant strengthening mechanisms, be they pragmatic or grammatical, are not limited to quantifiers such as *some*: they apply to other quantifiers (e.g., < *few, none* >), connectives (e.g., < *or, and* >), numerals (< *one, two, three …* >), verbs, adjectives, etc.

Despite the fact that the existence of three distinct readings is generally acknowledged in the literature,[4] experimental studies on scalar implicatures have typically used designs which are only able to distinguish two

[2] *Primary* and *secondary* scalar implicature are sometimes also referred to as *weak* and *strong* scalar implicature (e.g., Soames, 1982; Geurts, 2010; Bergen & Grodner 2012). Here we decided to avoid this terminology because *weak* is sometimes used to refer to the literal reading as opposed to the scalar implicature (*strong* reading).

[3] Other accounts have been proposed for SI: we refer the reader to Sauerland (2012) for a summary.

[4] Haida and Buccola (in press), however, argue that readings with the primary scalar implicature (and no secondary implicatures) do not exist for items like 'some'. On their view, a sentence such as (1) gives rise to the secondary scalar

readings (as we discuss below). In this paper, we have two related goals. First, by constructing an experimental design that is able to tease apart the three distinct readings, we will be in a position to address whether the derivation of the secondary implicature reading relies on the *Competence Assumption*, thereby contributing to the debate between pragmatic and grammatical approaches to SIs. Second, our design will allow us to reach a more fine-grained description of the pragmatic profile of different types of expressions, making it possible to compare various items which have been argued to give rise to scalar implicatures, including some controversial cases: the `exact' reading of numerals, the multiplicity inference triggered by plural morphology and the negative implication associated with *almost*.

Comparison of the two approaches

While the two approaches to SIs are quite different, they agree on a number of core facts: that three readings (literal reading, primary reading, secondary reading) can in principle exist for sentences containing scalar items, that alternatives play a critical role in the derivation of strengthened meanings, and that contextual factors determine which interpretation is accessed. However, they differ regarding the relations between the various readings, and importantly, regarding the role played by the listener's beliefs about the speaker's opinionatedness. Under the standard Neo-Gricean approach, the secondary reading will arise if and only if the listener assumes that the speaker is opinionated about the stronger alternative (*Competence Assumption*). Under the Grammatical approach, if the listener knows that the speaker is not opinionated about the stronger alternative, the secondary reading should not be derived either (since that would contradict the information that the speaker is not well-informed), but a primary reading may arise. But if the listener is merely *uncertain* as to whether the speaker is opinionated, she should treat the sentence as any other ambiguous sentence and both readings should be in principle possible – while on the standard neo-Gricean view, the secondary reading should not arise in such a case.

More recent game-theoretic pragmatic approaches (Franke, 2011; Goodman, & Stuhlmüller, 2013; Bergen et al., 2016, a.o.), which model participants as engaging in sophisticated higher-order reasoning about other participants' epistemic states, can also account for the derivation of the secondary reading when the *Competence Assumption* does not hold: even in a situation where the speaker is not well-informed, the speaker might reason that the listener *might* believe that the speaker is well-informed, and thereby choose her message as if the listener believed she were well-informed (avoiding, then, a *some* sentence when she does not believe 'not all'). The strengthening then involves the following steps: if the listener believes that the speaker is well-informed, they would derive a secondary *some but not all* reading. If the speaker believes that the listener may believe that they are well-informed, they will avoid using *some* when they do not believe *some but not all*. A listener may be aware of this reasoning on the part of the speaker, and may therefore associate uses of *some* with genuine *some but not all* reading. To put it differently, the risk that competence could be assumed even when it does not in fact hold, and not necessarily the *Competence Assumption* itself, could be sufficient to lead to the derivation of a secondary reading. Therefore, finding secondary readings even when the *Competence Assumption* does not hold would be evidence against the most traditional type of pragmatic neo-Gricean approaches, but would be compatible with these two types of models: a model where the secondary reading is derived as part of compositional semantics as in the grammatical approach, or a pragmatic model that posits sophisticated higher-order reasoning on the part of speakers and hearers.

Scalar diversity

The second issue we address in this paper is the question of how uniform and general the SI phenomenon is. As explained before, SI are not limited to quantifiers such as *some*. There is considerable variation however between the diverse alleged scalar items (see van Tiel, Miltenburg, Zevakhina, & Geurts, 2013): this challenges a uniform approach of the phenomenon and motivates a more systematic empirical investigation. Given that the experimental paradigm we present allows to test fine-grained properties of scalar items, we use it as a diagnostic tool of SIs, applying it to three debated cases of SI: numerals, plural morphology, and *almost*. As we will explain below, these three expressions may give rise to inferences that can, but don't have to, be analyzed as SIs: we will investigate whether they all give rise to three levels of interpretations in the same way as *some*.

Numerals

A sentence containing a numeral quantifier such as *two* in (2) can be taken to mean either that (2-a) 'at least' two students are home or that (2-c) 'exactly' two students are home. Under the traditional view (Horn, 1972), the interpretation of numeral quantifiers can be analyzed as involving SIs (for a summary of theoretical accounts, see Spector, 2013): the literal meaning of numerals is the 'at least' reading and the 'exactly' reading (2-c) results from a secondary implicature. There should also be a reading corresponding to just the primary implicature, described in (2-b).

---

implicature whenever the corresponding meaning is contextually relevant. As a result (1), according to them, triggers the inference that either it is not relevant whether all students are home, or not all students are home.

(2) Two students are home.

a. the speaker believes that at least two students are home (literal reading)

b. the speaker believes that at least two students are home and it is not the case that she believes that three or more than three students are home (primary implicature)

c. the speaker believes that exactly two students are home (secondary implicature)

This approach is not the dominant one today. A good deal of evidence suggest that there are substantial differences between numerals and standard scalar items. First, the distribution of the putative secondary reading described in (2-c) seems to differ from the strengthened meaning of standard scalar items (Horn, 1992; Breheny, 2008): the 'exactly' reading of numerals arises in more linguistic environments than the strengthened reading of standard scalar items. Second, there seems to be processing differences between numerals and standard scalar items, as shown using eye-tracking (Huang & Snedeker, 2009a) and dual-task paradigms (Marty, Chemla, & Spector, 2013). According to Marty et al., tapping memory resources has opposite effects on *some* and on numerals: under high cognitive load, participants derived fewer SIs for *some*, but more 'exact' readings for numerals. Third, acquisition studies have shown that children acquire the 'exactly' reading of numerals earlier than the strengthened meaning of other scalar items: Papafragou & Musolino (2003) found that 66% children aged of 5 year-old accessed the strengthened interpretation for numerals, whereas only 12,5% did for *some* (see also Huang & Snedeker, 2009). Note however that this result can be accommodated within a theory of numerals as SI, as these differences in children's performance might be due to a difference in the accessibility of different scales (Barner & Bachrach, 2010; Bale & Barner, 2013, a.o.).

Plural

The plural/singular distinction is the source of a long-standing debate in semantics. Intuitively, the meaning of plural marking seems to be 'strictly more than one' (Lasersohn, 1995), but in certain linguistic environments, for example under negation, it can be interpreted as meaning 'at least one'. For example, (3) will be considered false even if there is *only one* student in the room.

(3) There are no students in the room.

It has been proposed to treat the 'strictly more than one' component of the meaning of the plural morpheme is an implicature (Sauerland, Anderssen, & Yatsushiro, 2005; Spector, 2007; Zweig, 2008; Ivlieva, 2013). According to this view, the literal meaning of the plural morpheme is 'at least one'. Simplifying somewhat[5], the scale to consider is <plural, singular> (singular, meaning 'exactly one', asymmetrically entails plural). The implicature is thus 'at least two', and is expected to disappear in downward entailing environments, for example under negation. If the 'more than one' component arises as a scalar implicature (in a Neo-Gricean manner), we expect three distinct readings for a sentence such as (4):

(4) Students are in the room.

a. the speaker believes that at least one student is in the room (literal reading)

b. the speaker believes that at least one student is in the room, and it is not the case that she believes that just one student is in the room (primary implicature)

c. the speaker believes that several student are in the room (secondary implicature)

On some other accounts (Farkas & Swart, 2010, and, in a sense, Križ, 2017)[6], the plural morpheme is ambiguous between 'at least one' and 'at least two' interpretations, and speakers tend to favor the logically strongest reading, which happens to be the 'at least two' reading in simple sentences and the 'at least one' reading when the plural morpheme is in the scope of negation. On such an account, one would not expect to detect three distinct readings, but at most two, the weak 'at least one' reading and the strong 'at least two' reading. There are few experimental studies on this case, but Pearson, Khan, & Snedeker (2011) found that under certain circumstances, it was possible to cancel the 'more than one' meaning component. (This was also investigated in children (Tieu, Bill, Romoli, & Crain, 2014) and using priming studies (Chemla & Bott, 2015)).

Almost

Intuitively, the meaning of a sentence containing *almost X* as (5) seems to be 'almost X, but not X'. Most current theories analyze the 'but not X*'* meaning component of *almost* as part of its core semantics (Hitzeman, 1992; Horn, 2011; Kilbourn-Ceron, 2015). It has also been proposed to consider it as another case of SI (Sadock, 1981; Spector, 2014): under this view, (5-c) is obtained as a secondary implicature, based on the scale < almost X, X>. The reading (5-b) should also be available, corresponding to the primary implicature.

(5) Almost all students are home.

a. the speaker believes that all or not far from all students are home (literal reading)

---

[5] To be more precise, the plural can be viewed as a case of Higher Order Implicature. See Spector (2007) for details.

[6] Kriz (2017) proposes a trivalent analysis in which a sentence such as 'There are horses' is true if there are two or more horses, false if there are no horses, and neither true nor false if there is exactly one horse. Interestingly, this account can also be reformulated in terms of an ambiguity account, where the 'neither true nor false' case would correspond to a case where one reading if true and the other reading is false.

b. the speaker believes that all or not far from all students are home and it is not the case that she believes that all students are home (primary implicature)

c. the speaker believes that almost all but not all students are home (secondary implicature)

We use our paradigm to compare these two approaches: only under the SI view should we observe three interpretations.

The experimental study of SI

Truth value judgment tasks

Diverse methodologies have been used to investigate SI, both in adults, with truth value judgment tasks, response-time studies (Bott & Noveck, 2004; Noveck & Posada, 2003), self-paced reading (Breheny, Katsos, & Williams, 2006; Bergen & Grodner, 2012), mouse-tracking (Tomlinson, Bott, & Bailey, 2011), eye-tracking (Huang & Snedeker, 2009; Grodner, Klein, Carbary, & Tanenhaus, 2010), ERPs (Politzer-Ahles, 2013; Hunt, Politzer-Ahles, Gibson, Minai, & Fiorentino, 2013; Spychalska, Kontinen, & Werning, 2016), dual-task paradigms (De Neys & Schaeken, 2007; Dieussaert, Verkerk, Gillard, & Schaeken, 2011; Marty & Chemla, 2013; Marty, Chemla, & Spector, 2013) and in children (Smith, 1980; Noveck, 2001; Papafragou & Musolino, 2003, among others). In classical sentence-verification tasks (Bott & Noveck, 2004) (based on paradigms implemented by Rips, 1975, and Smith, 1980), participants are presented with underinformative sentences, for example "Some elephants are mammals*",* and have to decide whether the sentence is true or false. If participants access the literal meaning of the sentence, they judge it as true, whereas if they access the implicated meaning, they judge it as false. This rationale allows to establish the existence of two distinct levels of interpretations: the mere literal meaning (LIT) and an interpretation with implicature (SI). However, in such a paradigm, the PSI and the SSI are lumped together: to judge the sentence as false, it is sufficient to derive the PSI, as it is false that the speaker does not hold the belief that all elephants are mammals if it is common knowledge that all elephants are mammals. To our knowledge, no study directly tests the existence of three distinct interpretations.

We now turn to three studies that investigate the effect of the epistemic state of the speaker on SI computation and thus indirectly address our issue. First, Bergen & Grodner (2012) (following a previous experiment by Breheny, Katsos, & Williams (2006)), explicitly discuss the question of PSI and SSI (which they refer respectively to Weak SI and Strong SI) using a self-paced paradigm. The goal of their experiment is to investigate the effect of the speaker's level of expertise on the derivation and timing of SI. Each trial consists of three sentences: a context sentence, a trigger sentence and a continuation sentence. The context sentence determines the information level of the speaker (for example: Full-knowledge – "I meticulously compiled the investment report". vs. Partial-knowledge – "I skimmed the investment report."). The target sentence is then: "Some of the real estate investments lost money". There is then the continuation sentence: "The rest were successful despite the recent economic downturn." The dependent variable is the reading time on *some* and on *the rest*: the derivation of an implicature early on (i.e. when reading *some*) is indicated by a longer reading time on *some* and a shorter reading time on *the rest* (the presupposition of *the rest*, i.e. that not all investments lost money, is automatically satisfied if the 'not all' implicature has been computed when encountering *some*, and creates an additional cost when this is not so). The expectation is that there should be more SI in the Full-Knowledge condition than in the Partial-Knowledge condition, and this is indeed what the authors found. At first sight, this study could seem to establish the existence of SSI *per se:* the presupposition of *the rest* (that not all investments lost money) is automatically satisfied by previous discourse only if *some* has been enriched to 'some but not all'. However, this result does not directly establish the existence of two distinct types of implicatures (PSIs and SSIs), for two reasons. First, if the only possible readings were the literal reading and the secondary reading, the results would be expected: in the case of two readings that are ordered in terms of logical strength, the disambiguation process should give more plausibility to the weaker (literal) reading when the speaker is not presented as being well informed, and to the stronger (SI) reading if the speaker is supposed to be well informed. Second, if, conversely, only the literal and PSI readings exist, then in a context where the speaker is known to be well informed, the PSI is contextually equivalent to the SSI, which may explain the facilitation found on *the rest*. Overall, this paradigm only distinguishes between two possibilities (fast and slow responses on *the rest*) by design, and therefore cannot offer direct evidence for a three way distinction of interpretations.

Second, Breheny, Ferguson, & Katsos (2013) present a sophisticated study based on a visual-world paradigm. In their experiment, participants have to listen to a speaker (a confederate) who is describing a video. The speaker sees a video in which an agent puts one or several objects into two boxes. She then picks a sentence to describe what she saw to the listener, who sees a still image of the last frame of the video so knows which objects were put in which boxes. In some conditions, the speaker sees the whole video (*fully-informed speaker*), while in others she sees only half of it (*under-informed speaker*). The level of information of the speaker is common knowledge between speakers and listeners. In the critical condition, the agent in the video puts a spoon and nothing else in box A, then a spoon in box B, and then a fork in box B (the end-result is visible to the hearer). If she is fully-informed and says, e.g., "The woman put a spoon into box A and a spoon into box B", the listener can infer right after the first occurrence of 'into' that the next words will be `box A', by deriving a quantity implicature: since the speaker said 'a spoon into' and not 'a spoon and a fork into', she must be talking about box A (which has only a spoon). In contrast, when the speaker is not presented as fully-informed (she didn't see the whole video, and the listener knows it), the listener cannot be sure that the speaker is talking about box A after hearing the first occurrence of 'into'. What was found is that when (and only when) the speaker is presented as knowledgeable, listeners look at the target before it is mentioned (box A in the above case). These results show that listeners take into account what they know about speakers' epistemic state when they interpret utterances, but here again, the experimental design does not tease apart primary and secondary implicatures, simply because like the previous one it only distinguishes between two possibilities. These results are compatible with the sentence being simply ambiguous between a non-exhaustive reading ('a spoon and maybe something else') and an exhaustive one ('a spoon and nothing else'). Listeners would not disambiguate in favor of the secondary reading when the speaker is presented as under-informed about the stronger alternative[7]: in this case, it is irrational to assume that she intends to convey the secondary reading, as it implies that she is opinionated. These results thus do not show that a primary implicature is derived in the absence of a secondary implicature when the speaker is presented as not being opinionated: primary implicatures *per se* are simply not tested.

Another relevant study is Hochstein, Bale, & Barner's (2017), which addresses the computation of SIs in autistic children. They use a Partial-Knowledge task as a test of Epistemic Reasoning (Experiment 3), which has some similarities with the paradigm we used. The principle is the following: there are three boxes. A character opens two of the three boxes, making it visible to the participant that both boxes contain, for instance, strawberries. In the *full-knowledge* condition, the character opens the third box and looks into it, but the participant cannot see the content of the box. In the *partial-knowledge* condition, the character does not open the third box. Then the character says: "Some of the boxes have strawberries"[8]. The question asked is: "Do you think there are strawberries in this box?" (where "this box" refers to the third box, whose content in both conditions is not visible to the participant). The *full-knowledge* condition makes it possible to distinguish the literal reading (*I don't know* answers) from SI (*no* answers) (the classical condition to test SI). In the *partial-knowledge* condition, the expected answer is *I don't know*: subjects are not expected to draw the secondary implicature, as the speaker is not in a position to know whether all of the boxes have strawberries. Note that this is just a consequence of the situation described: the speaker cannot have meant the stronger reading, given that we know that he is not in a position to have the relevant knowledge. The results show that in the *partial-knowledge* condition, autistic participants (but not controls) tend to answer *no* instead of *I don't know*, which is interpreted by the authors as showing that they are more likely to compute SI in incorrect contexts (i.e., without epistemic justification). It is important to note that in this experiment, the subjects who say *no* not only may have computed an SSI, but they are in any case ignoring the relevant features of the situation (i.e., the fact that speaker *cannot* know that the SI is true). While this may suggest that the secondary interpretation is a possible reading for autistic people[9], this design, like the one used in Breheny et al. (2013), still does not establish the existence of 3 distinct readings, neither in the general population nor in the autistic population.

To sum up, these three studies distinguish between the literal reading and the reading with a scalar implicature, and provide some information about the effect of the perceived epistemic state of the speaker on interpretation, but they do not demonstrate the availability of the three interpretations we were interested in. At least to our knowledge, it has never been shown straightforwardly that the primary and the secondary implicatures were distinct. The paradigm we now present is close to Breheny et al.'s (2013) and Hochstein et al.'s (2017), in that it relies on the manipulation of the speaker's information level; however, it differs from it regarding the question asked to the participant.

[7] The complete design, which we do not describe here, ensures that when the speaker does not see the full video, the listener knows that the speaker does not know whether the agent put one or two different objects in the second box.

[8] Numerals and *all* were also tested (in the case of *all,* only in the *full knowledge* condition). We focus here on the aspects of the design relevant to us.

[9] There is however another possible interpretation: it might be that autistic children only derive the PSI but quite generally act as if the speaker were well informed, ignoring the critical feature of the *partial knowledge* condition. In such a case, the PSI would be, for them, contextually equivalent to the SSI. To rule out such an interpretation, one would need to test the control sentence with *all* in the *partial knowledge* condition. In the normal population, the conclusion should be that the character is just guessing and is therefore not reliable. If, however, autistic children act as if the speaker is well informed, they should answer that there are strawberries in the third box. Hochstein et al.'s design includes an intermediate step where the participant is asked whether the character knows what is in the third box, and the question is repeated until the participant uses the correct answer (*no* in the *partial knowledge* condition). The paper does not indicate how many trials autistic children need to correctly answer this question. The fact that they end up answering correctly is in this case a consequence of the design (since they are not allowed to proceed until they provide the correct answer), and does not provide strong evidence that they understand that the character is ignorant about the third box.

Experiment 1: Primary and secondary implicatures for different scalar items

Goal

The two goals of this experiment were a) to test the role of the *Competence Assumption* in the derivation of the SSI reading, and b) to investigate three debated cases of SI: the exact reading of numerals (NUMERALS), the plural reading of the plural morpheme (PLURAL) and the negative inference associated with the word *almost* (ALMOST), comparing their behavior to the "standard" case of SI, SOME. To do so, we needed an experimental design able to distinguish the three interpretations.

Design

The experiment was a truth-value judgment task. Participants were presented with a display representing the information available to two players, Peter and Mary, involved in a card game, as illustrated in Figure 1. The display showed two versions of a set of 8 cards, meant to correspond to two different information levels: one of the players (Mary in Figure 1) was fully informed and able to see all of the cards (*Knowledgeable Player*), while the other player (Peter in Figure 1) had missing information, represented by cards with a question mark (*Ignorant Player*). In no cases was there a mismatch between the cards visible to both players: the only difference was the presence of hidden cards for one but not the other. A sentence was displayed on the screen, and attributed to one of the player. The speaker could thus be either fully informed (*Knowledgeable Speaker*) or only partially informed (*Ignorant Speaker*) (Mary in Figure 1)[10]. Participants were asked to judge whether the speaker Mary(/Peter) could or could not have said the sentence, according to her informational state. Two answers were possible: "Mary(/Peter) can say that" or "Mary(/Peter) cannot say that". Importantly, with this type of judgment, there were two reasons for responding negatively: (i) the speaker knows that the sentence is false (it conflicts with her information), and (ii) the speaker does not have enough information to know whether the sentence is true or false. For the sake of brevity, we will use *yes* and *no* to refer to these answers.

Figure 1. Example of the display presented to the participants (*Ignorant Speaker*).

The sets of cards were chosen to construct four conditions in combination with the target sentences (see Table 2). We refer to the conditions according to the reading(s) the set makes true: either no reading ('false' condition), the literal reading only ('literal' condition), the literal reading and the primary implicature but not the secondary implicature ('primary' condition), or the three readings ('secondary' condition). 'False' and 'secondary' conditions thus correspond to 'no' and 'yes' controls, respectively. Our predictions were as follows:

- In the 'literal' condition, the speaker is *knowledgeable*: she knows that the stronger scalar alternative is true. (This is the classical case used to test implicatures, e.g., in Bott & Noveck, 2004).

[10] Half of the information presented at each trial was thus in principle useless: the cards seen by the other player were supposed not to influence the answer. We presented the two sets of cards to ensure that the actual situation (as shown by the fully informed player) was available and to control that participants actually based their decision on the speaker's knowledge, beyond what the actual world was shown to be like.

- If the participant accesses LIT, she will answer *yes*.
- If she accesses the PSI or the SSI, she will answer *no*.

- In the ‘primary’ condition, the speaker is *ignorant*: she knows that the literal meaning is true but does not know whether the stronger scalar alternative is true. This new case distinguishes between PSI and SSI.
  - If the participant accesses LIT or PSI, she will answer *yes*.
  - If she accesses the SSI, she will answer *no*.

Table 2. Speaker’s cards for the four scalar items tested.

Scalar Item		*some*	numeral	plural	*almost*	
					almost all	*almost no*
Sentence		*Some of the cards are hearts.*	*Two cards are hearts.*	*Some cards are hearts.*	*Almost all cards are hearts.*	*Almost no card cards is a heart.*
Condition	false (control no)	♤♤♤♤ ♤♤♤♤	♤♤♤♤ ♤♤♤♤	♤♤♤♤ ♤♤♤♤	♤♤♤♤ ♤♤♤♤	♥♥♥♥ ♥♥♥♥
	literal	♥♥♥♥ ♥♥♥♥	♥♥♥♥ ♤♤♤♤	♥♤♤♤ ♤♤♤♤	♥♥♥♥ ♥♥♥♥	♤♤♤♤ ♤♤♤♤
	primary	♥♥♥♥ ????	♥♥♤♤ ????	♥ ♤♤♤ ????	♥♥♥♥ ♥♥♥?	♤♤♤♤ ♤♤♤?
	secondary (control yes)	♥♥♥♥ ♤♤♤♤	♥♥♤♤ ♤♤♤♤	♥♥♥♥ ♤♤♤♤	♥♥♥♥ ♥♥♥♤	♥♤♤♤ ♤♤♤♤

These 4 conditions (‘false’, ‘literal’, ‘primary’ and ‘secondary’) were further divided into 7 conditions:

- The ‘false’ condition was subdivided into:
  - ‘false-kn’: Expected answer: *no* because the sentence is false and the speaker knows it.
  - ‘false-ign’: Expected answer: *no* because the speaker does not have enough information, itself divided into:
    - ‘false-ign-true’: the sentence is plainly true in the actual world (under the other player’s view).
    - ‘false-ign-false’: the sentence is plainly false in the actual world (under the other player’s view).
- The ‘primary’ condition was subdivided into:
  - ‘primary-ign-true’: all readings are true in the actual world.
  - ‘primary-ign-ambig’: the reading with an implicature is false in the actual world (but the literal reading is true).

These subdivisions allowed us to control that the actual world did not influence the answers and that participants interpreted cards with question marks as representing ignorance, rather than another type of card, showing that participants really based their answers on the knowledge of the alleged speaker. Target cases (‘literal’ and ‘primary’) represented 33% of trials for each type of target sentence. There was always the same number of *yes* and *no* controls. Further details are given in Appendix B.

Materials

Sentences

Four scalar items were tested: SOME, NUMERALS, PLURAL and ALMOST. Sentences were always of the form “X card(s) are/is a Y”, with X in “Some of the”, “Two”, “Some”[11], “Almost all”, “Almost no”, “All” or “No” and Y in “heart(s)”, “diamond(s)”, “spades(s)”, “club(s)” (see Table 3). Controls with “No” were included to check that participants correctly understood the meaning of question marks, and did not interpret them as representing another type of cards (e.g., a card which was not a heart): when no heart is visible and some of the cards are hidden, the participant will answer *yes* (“Peter can say that”) only if she incorrectly understands that the hidden card is not a heart, whereas she will answer *no* if she correctly understands that the hidden card *might* be a heart.

[11] Trials with *some* rather than *some of the* tested the interpretation of plural morphology, not the *some-but-not-all* implicature.

Table 3. Target and control sentences

Target sentences	
Some	Some of the cards are hearts[12].
Almost X	Almost all cards are hearts. Almost no card is a heart.
Numerals	Two cards are hearts.
Plural	Some cards are hearts.
Control sentences	
No	No card is a heart.
All	All of the cards are hearts.

Pictures
As illustrated in Figure 1, the pictures consisted of two sets of 8 cards, one for each player, although only the set seen by the speaker was supposed to be relevant for the judgment task. In half of the trials, Mary was the speaker, and in the other half, Peter was the speaker. Mary's set of cards was always on the left, and Peter's on the right. The rank (from 1 to King) of the cards was chosen randomly, in the same way for all participants.
Procedure
The experiment was implemented online on Alex Drummond's Ibex Farm. After having given their consent to participate in the experiment, participants were given instructions (see Appendix D). There was first a training phase consisting of 14 non-controversial items with feedback, followed by the main experimental phase consisting of 288 trials without feedback. There was no explicit transition between the two phases but the first four sentences of the experimental phase were taken from the training phase, in order to get the subjects used to the no-feedback part of the experiment, and were removed from the analysis. At the end of the experiment, there was a short questionnaire asking for age, sex, native language, kind of device used to answer, and Mechanical Turk Worker ID.

Participants
60 participants were recruited via Mechanical Turk. They were paid $2 for their participation. 59 of them completed the task. We removed from the analysis 1 participant who did not declare English as her native language, 5 that made more than 35% (mean-2*standard deviation) errors on controls, and 1 who made more than 30% errors on NO-controls. We thus present the results for 52 participants (36 females, 16 males, mean age: 41,8, from 20 to 62 year old).

Results
Data analyses were conducted using R. We used binomial linear mixed effects model, built with a maximal random effect structure based on subjects and items as random variables, although we sometimes had to step back to random-intercepts-only models when the model failed to converge with the full random-effects specification (following Barr et al., 2013). The data and analysis scripts for all the experiments we report on are available on-line: http://semanticsarchive.net/Archive/2Q1MWRlM/.

Analysis of responses
Data treatment: we removed the trials that were below 200 ms and above 20000 ms (less than 1% of the data). We then removed, for each participant, the trials that were above and below 2 standard deviations from the mean, keeping 94,9% of the data.

The proportion of errors on controls was very low (all controls: 2,4%; 'false' condition: 2,1%; 'secondary' condition: 2,7%). Figure 2 shows the overall proportion of *yes* answers in the 4 conditions for target sentences. Under the assumption that participants were coherent in their interpretation, we can infer the rates of the different interpretations: 37% of literal interpretations (*yes* in the 'literal' condition), 45% of SSI interpretations (*no* in the 'primary' condition) and 18% of PSI interpretations (the difference between *yes* responses in the 'literal' and the 'primary' condition, only the latter one involving PSI).

Figure 3 presents the results split by sub-conditions. We find no significant difference between the different instantiations of the 'primary' conditions, comparing the conditions in which the sentence is true under all interpretations in the actual world ('primary-ign-true') and those where it is true only under the literal

[12] The partitive construction (*Some of the cards are hearts* vs. *Some cards are hearts*) is known to favor SI computation (Grodner, Klein, Carbary, & Tanenhaus, 2010). Given that the purpose of our experiment was to study PSI and SSI, we chose to use the partitive construction.

interpretation ('primary-ign-ambig'): the 'actual world' is not found to affect the answers of the participants, which shows that participants are indeed judging the sentences depending on the speaker's knowledge state and not on the actual world.[13]

Figure 2. Proportion of *yes* answers for all scalar items (4 conditions)

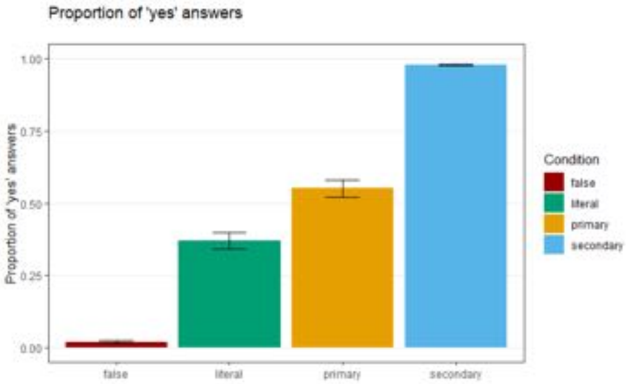
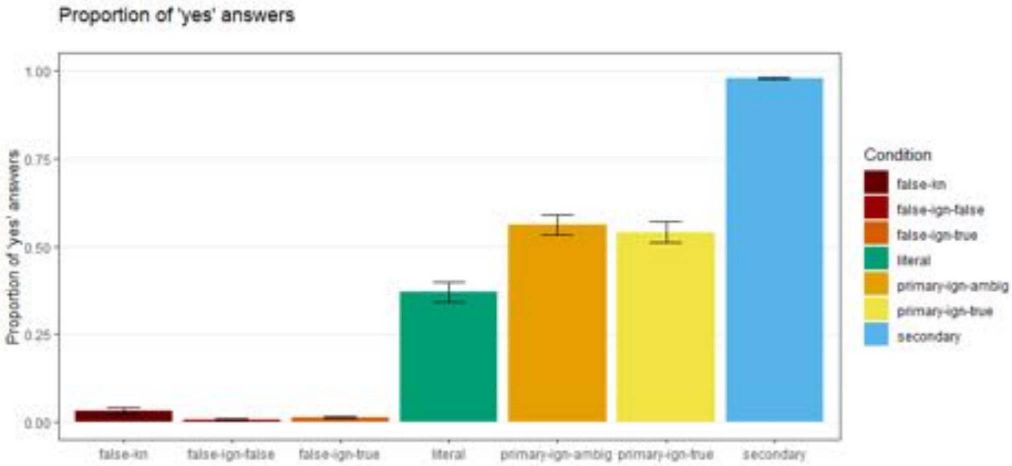


Figure 3. Proportion of *yes* answers for all scalar items (7 conditions)



Existence of three interpretations

We fitted a linear mixed model to predict Answer (*yes* vs. *no*), using Condition as fixed variable and Subject and Item as random factors. We first ran an omnibus test for all scalar items, and then one for each scalar item separately. Three contrasts enabled us to distinguish between the three interpretations.

First, we compared 'false' to 'literal' in order to detect LIT. There was a significant difference between the rate of *yes* answers between the two cases (χ^2 *(1) = 67, p < 4.10*$^{-16}$). We then compared 'literal' to 'primary' in order to detect the PSI (χ^2 *(1) = 13, p < .001*). We finally compared 'primary' to 'secondary' in order to detect the SSI (χ^2 *(1) = 70, p < 3.10*$^{-16}$).

Results for each scalar item are shown in Figure 4, and corresponding analyses are given in Table 4. For SOME and ALMOST, we detected three interpretations: the differences are significant for the three tests. For TWO and PLURAL we detected LIT and SSI, but not PSI: the differences between 'literal' and 'primary' are not significant (TWO: χ^2 *(1) = 0.62, p = .43*; PLURAL: χ^2 *(1) = 0.35, p = .55*).

In order to confirm the difference of behavior between different scalar items, we also tested the interactions between a) the type of scalar items and b) the two conditions that are relevant to testing the PSI reading, i.e. the 'literal' and 'primary' conditions. Results of interaction tests are summed up in Table 5. As expected, the

[13] Following a reviewer's suggestion, we ran posthoc analyses where we compare the two sub-conditions for the 'primary' and 'false' conditions separately for each item. After correction for multiple comparisons, only one of 16 comparisons turned out to be significant: the comparison between the sub-conditions where the expected answer was *no* because the speaker does not have enough information, and the sentence is plainly true in the actual world ('false-ign-true'), and the one the expected answer was *no* because the speaker does not have enough information, and the sentence is plainly false in the actual world ('false-ign-false'). None of these comparisons turn out significant for Experiment 2. The full results are reported in Appendix E.

interaction is significant in all cases but two: SOME and ALMOST do not differ significantly and TWO and PLURAL do not either.

To address a reviewer's remark, we also ran posthoc analyses where we restricted the comparisons to sub-conditions such that the 'actual' world is the same in the two conditions being compared, i.e. 'literal' vs. 'primary-ign-false' (comparison between the 'literal' condition with the 'primary' sub-condition where the actual world is as in the 'literal' condition), and 'primary-ign-true' vs. 'secondary' (comparison between the 'primary' sub-condition where the SSI is true in the actual world and the 'secondary' condition where the speaker knows that the SSI is true). Results are summarized in Table 6. The only difference was that in the case of SOME, where the difference did not reach significance ($\chi^2(1) = 3.54$, $p = 0.059$) (but the very same comparison turns out significant in our second experiment).

Figure 4. Proportion of *yes* answers by condition and scalar item (4 conditions)

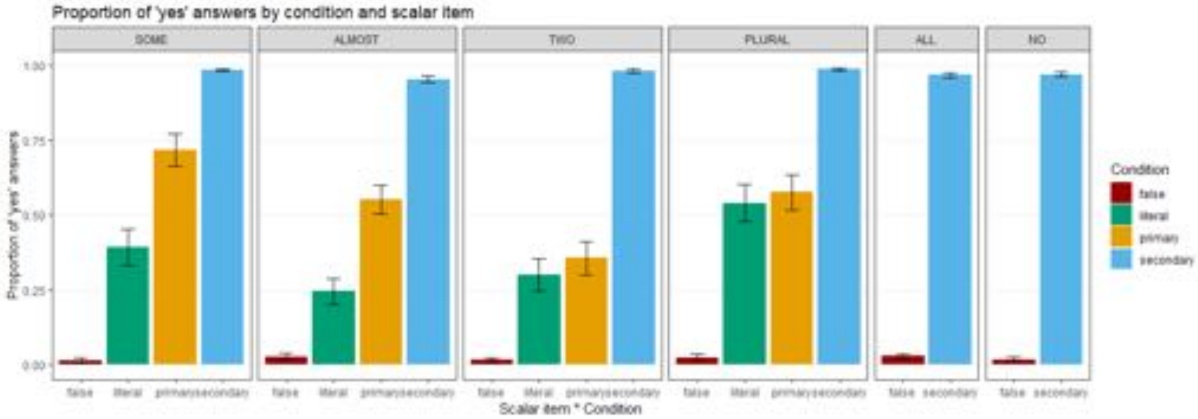


Table 4. Differences between conditions revealing the existence or absence of each of the three readings, for each of the four potential scalar items

	LIT: 'false' vs. 'literal'	**PSI:** 'literal' vs. 'primary'	**SSI:** 'primary' vs. 'secondary'
SOME	*$\chi^2(1) = 10,9$, $p = 0.001$ ****	*$\chi^2(1) = 18,9$, $p = 1.4e-05$****	*FTC with full specification* *$\chi^2(1) = 55$, $p < 1.6e-13$ ****
ALMOST	*$\chi^2(1) = 10.9$, $p = .00092$ ****	*$\chi^2(1) = 21.5$, $p = 3.5e-06$ ****	*$\chi^2(1) = 39.1$, $p = 4e-10$ ****
TWO	*FTC with full specification* *$\chi^2(1) = 307$, $p < 2.2e-16$ ****	*$\chi^2(1) = 0.62$, $p = .43$*	*FTC with full specification* *$\chi^2(1) = 750$, $p < 2.2e-16$ ****
PLURAL	*$\chi^2(1) = 14$, $p = .00018$ ****	*$\chi^2(1) = 0.35$, $p = .55$*	*$\chi^2(1) = 14,8$, $p = .00012$ ****

Table 5. Two-by-two comparisons of scalar items concerning the existence and robustness of the PSI interpretation (as measured by the difference between the 'literal' vs. 'primary' conditions). Formally, each cell reports the interaction between scalar items taken two by two (e.g, SOME vs. ALMOST) and condition (always 'literal' vs. 'primary').

'literal' vs. 'primary'	ALMOST	TWO	PLURAL
SOME	*$\chi^2(1) = 0.31$, $p = .58$*	*FTC with full specification* *$\chi^2(1) = 31.2$, $p = 2.4e-08$ ****	*$\chi^2(1) = 39.5$, $p = 3.3e-10$ ****
ALMOST	-	*$\chi^2(1) = 13.7$, $p = .0002$ ****	*$\chi^2(1) = 18$, $p = 2.3e-05$ ****
TWO	-	-	*FTC with full specification* *$\chi^2(1) = 0.27$, $p = .60$*

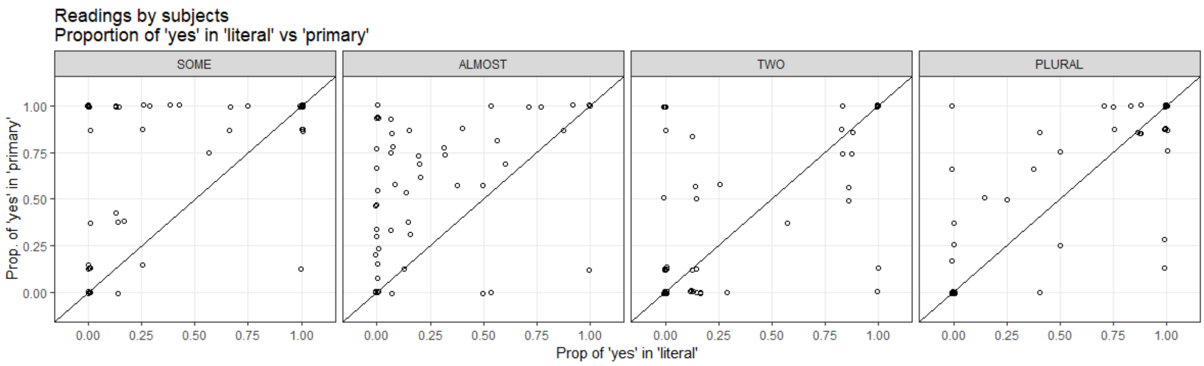
Table 6. Differences between conditions for each of the four potential scalar items (in rows), restricted to conditions such that the actual world is the same in the two conditions being compared.

	LIT (previously 'literal' vs. 'false') 'literal' vs. 'false-kn'	**PSI** (previously 'literal' vs. 'primary') 'literal' vs. 'primary-ign-ambig'	**SSI** (previously 'primary' vs. 'secondary') 'primary-ign-true' vs. 'secondary'
SOME	$\chi^2(1) = 15.5, p = 8.3e\text{-}05$***	$\chi^2(1) = 22.3, p = 2.37e\text{-}06$***	$\chi^2(1) = 3.54, p = 0.059$
ALMOST	$\chi^2(1) = 14.6, p = 0.00013$***	$\chi^2(1) = 19.2, p = 1.2e\text{-}05$***	$\chi^2(1) = 32.5, p = 1.18e\text{-}08$***
TWO	*FTC with full specification* $\chi^2(1) = 44.7, p = 2.2e\text{-}11$***	$\chi^2(1) = 1.07, p = 0.3017$	*FTC with full specification* $\chi^2(1) = 47.5, p = 5.36e\text{-}12$***
PLURAL	*FTC with full specification* $\chi^2(1) = 49.0, p = 2.5e\text{-}12$***	*FTC with full specification* $\chi^2(1) = 1.15, p = 0.2827$	*FTC with full specification* $\chi^2(1) = 38.0, p = 6.97e\text{-}10$***

Results by subjects

Figure 5 shows the responses of each participant combining the two target conditions 'literal' and 'primary', for each scalar item tested. Each data point corresponds to a participant (a jitter was added in order to make the results readable). Participants who consistently access LIT (answering *yes* to both 'literal' and 'primary' conditions) are in the top right corner. Participants who access the PSI (answering *no* to 'literal' and *yes* to the 'primary' condition) are in the top left corner. Participants who access the SSI (answering *no* in both cases) are in the bottom left corner. The fourth corner would show inconsistent behavior (answering *yes* more often to a condition ('literal') which makes strictly fewer readings true) and, as can be seen, is not much populated. Coherent with the results described above, the three relevant corners are populated for SOME and ALMOST, while only those corresponding to the LIT and SSI clearly emerge for PLURALS and NUMERALS.

Figure 5. Results by subject for the different scalar items



Arguably, an issue with the design of Experiment 1 was that there was no 'secondary' control condition in which the participant had to answer *yes* with a speaker who does not see all the cards: *no* responses in the 'primary' condition could have been driven by the superficial fact that the speaker was not fully knowledgeable. Experiment 2 was designed to address this concern.

Experiment 2

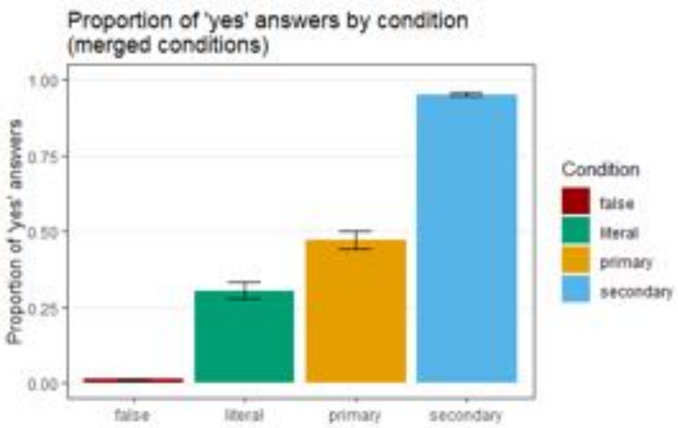
Experiment 2 was identical to Experiment 1. The only difference was that there was another 'secondary' control condition where the subject had to answer *yes* with a speaker who does not see all the cards, so as to address the problem presented above. In this new condition, the fictitious speaker *knew* that the stronger alternative (e.g., *all*) was false, despite not being *fully knowledgeable,* and so could use the sentence under all its interpretations. For instance, the speaker does not see all of the cards, but among the cards she sees, some are hearts and some are not hearts. To sum up, we found that participants behaved in the same way as before, and had no difficulty answering *yes* with a partly ignorant speaker when appropriate, thus offering a replication of the main findings of Experiment 1.

Method and materials
The material was the same as in Experiment 1, except that we added a version of the 'secondary' condition (*yes* control) in which the speaker knew that the secondary implicature was false but was not *fully knowledgeable* because she did not see all the cards (*Ignorant Player*). We refer to this new condition as 'secondary-ign' as opposed to 'secondary-kn'. In this condition, the speaker is *opinionated* about the stronger alternative even though she is *not fully informed* about all the cards (which is why we still employ the label 'ign'). For sentences testing SOME and PLURAL, this condition corresponded to the cards [♥♥♥♤♤???]. For sentences testing TWO, ALMOST, ALL and NO, such a condition is impossible to implement, as one needs to have all the information to answer *yes*. In order to increase the proportion of conditions in which the expected answer was *yes* despite ignorance of the speaker, we thus introduced sentences with FIRST and LAST (e.g. "The first card is a heart", with [♥♥♤♤????]). We also added control trials corresponding to 'secondary-kn', 'false-ign-false', and 'false-kn'. Experiment 2 consisted in 320 trials (further details are given in Appendix B). Overall, 20% of 'secondary' were 'secondary-ign'. The proportion of controls ('secondary' and 'false') and target cases was the same as in Experiment 1. The rest of the procedure was exactly the same as in Experiment 1.

Participants
60 participants were recruited via Mechanical Turk. 59 of them completed the task. We removed from the analysis 2 participants whose native language was not English, 6 that made more than 46% (m-2sd) of errors on controls, and 6 who made more than 30% of errors on NO-control. We thus present results for 45 participants (30 females, 15 males; mean age: 35,5, from 19 to 61 y.o.).

Results
Analysis of responses
We removed the trials that were below 200 ms and above 20000 ms (1,4% of the data). We then removed, for each participant, the trials that were above and below 2 standard deviations from the mean, keeping 94,2% of the data. Figure 6 shows the proportion of *yes* in the 4 conditions, Figure 7 according to the detailed conditions. Figure 8 shows the results for each scalar item. As in the first experiment, the mean of errors on controls was low (all controls: 3,1% ; 'false': 1,1% ('false-kn': 1,2%, 'false-ign-true': 1,1%, 'false-ign-false': 1,0%); 'secondary': 5,2% ('secondary-kn': 4,4 %,; 'secondary-ign': 8,4%). The mean rate of *yes* answers in 'literal', corresponding to LIT alone, was 30% (vs. 37% in Experiment 1)). The mean rate of *no* answers in 'primary', corresponding to SSI alone, was 53% (vs. 44% in Experiment 1). PSI thus represented 17% of the readings. As in Experiment 1, there was no significant difference between the different instantiations of the 'primary' and 'false' sub-conditions. Most importantly, there was no difference involved in the new distinction for the 'secondary' condition ($\chi^2(1) = 0.18$, $p = .64$).[14] We run the same statistical tests as for Experiment 1. They are summed up in Table 7 and 8.

Figure 6. Proportion of *yes* answers (4 conditions)



[14] Following a reviewer's suggestion, we also compared the relevant sub-conditions for each item separately. Only one of these 16 posthoc comparisons turns out significant before correcting for multiple comparisons, and significance is not preserved after correction for multiple comparisons. See Appendix E for detailed results.

Figure 7. Proportion of *yes* answers (detailed conditions)

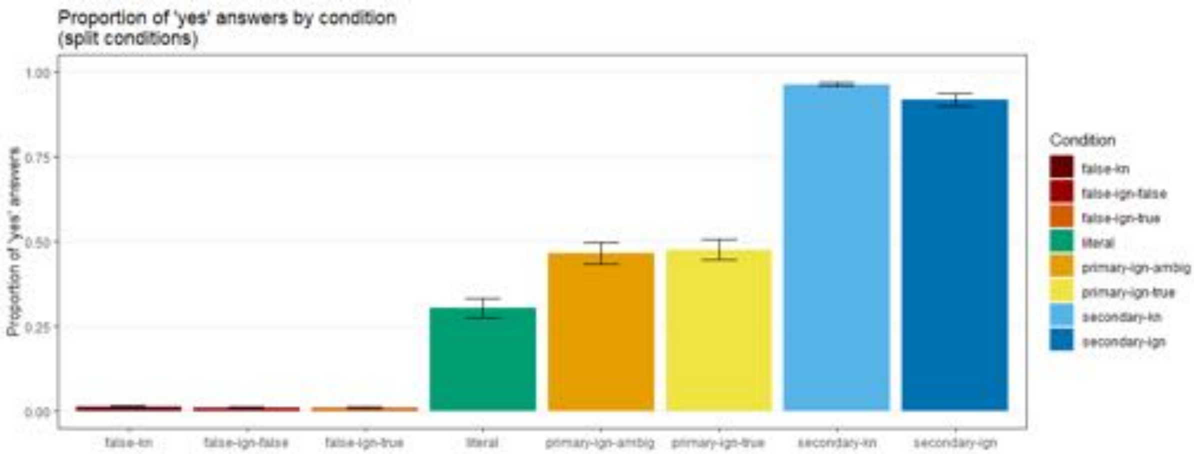


Figure 8. Proportion of *yes* by condition and scalar item

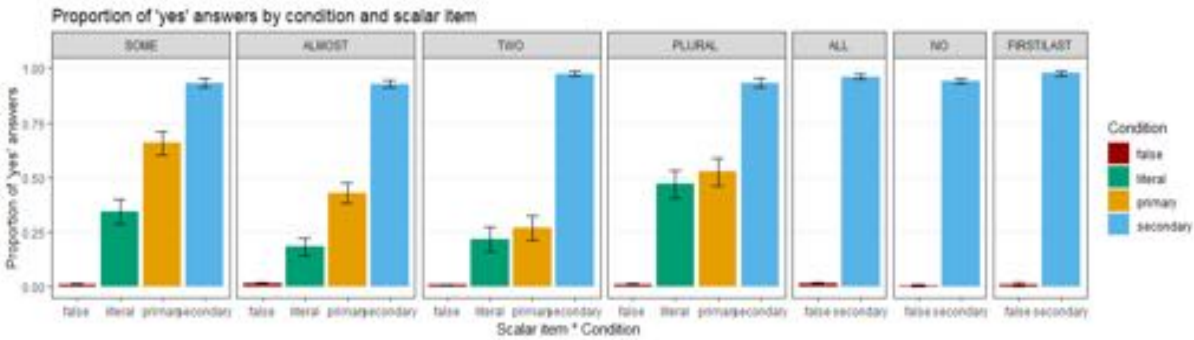


Table 7. Differences between conditions revealing the existence or absence of each of the three readings (in columns), for each of the four potential scalar items (in rows)

	LIT: 'false' vs. 'literal'	**PSI:** 'literal' vs. 'primary'	**SSI:** 'primary' vs. 'secondary'
SOME	$\chi^2(1) = 19$, $p = 1.5e\text{-}05$ ***	$\chi^2(1) = 22$, $p = 2.8e\text{-}06$ ***	$\chi^2(1) = 28$, $p = 1.1e\text{-}7$ ***
ALMOST	$\chi^2(1) = 8.8$, $p = .003$ ***	$\chi^2(1) = 18$, $p = 2.3e\text{-}05$ ***	$\chi^2(1) = 45$, $p = 2e\text{-}11$ ***
TWO	*FTC with full specification* $\chi^2(1) = 53.2$, $p = 3e\text{-}13$ ***	$\chi^2(1) = 0$, $p = 1$	*FTC with full specification* $\chi^2(1) = 69$, $p < 2.2e\text{-}16$ ***
PLURAL	$\chi^2(1) = 19$, $p = 1.3e\text{-}05$ ***	$\chi^2(1) = 2.4$, $p = 0.12$	*FTC with full specification* $\chi^2(1) = 46$, $p = 1e\text{-}11$ ***

Table 8. Two-by-two comparisons of scalar items concerning the existence and robustness of the PSI interpretation (as measured by the difference between the 'literal' vs. 'primary' conditions).

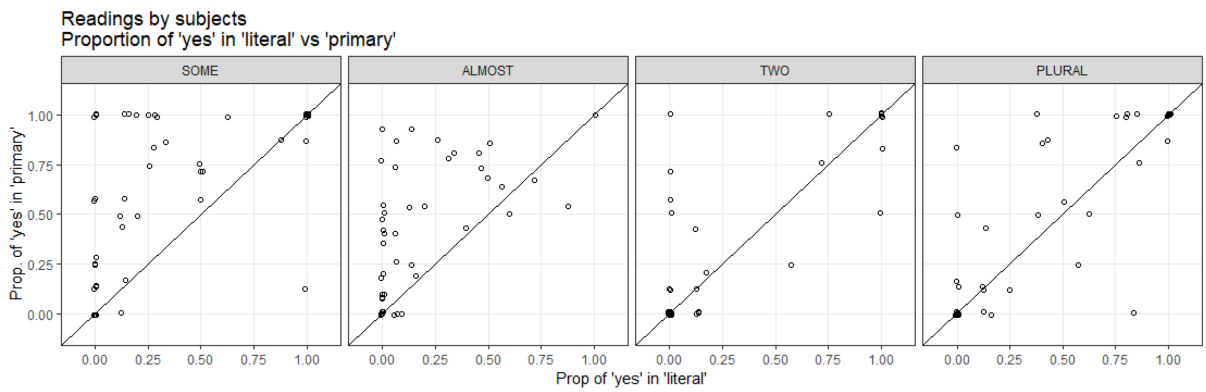
'literal' vs. 'primary'	ALMOST	TWO	PLURAL
SOME	*$\chi^2(1) = 0.08, p = .77$*	*$\chi^2(1) = 21, p = 4.5e\text{-}06$ ****	*$\chi^2(1) = 29, p = 7.9e\text{-}08$ ****
ALMOST	-	*$\chi^2(1) = 8.2, p = .004$ ***	*$\chi^2(1) = 9.4, p = .002$ ***
TWO	-	-	*$\chi^2(1) = 0.04, p = .84$*

To address a reviewer's remark, we again ran posthoc analyses where we restricted the comparisons to sub-conditions such that the 'actual' world is the same in the two conditions being compared. Detailed results are given in Table 9: they are qualitatively similar, except that one more comparison turns out significant ('literal' vs. 'primary-ign-ambig' in the case of *two)*.

Table 9. Differences between conditions, for each of the four potential scalar items (in rows), where the conditions being compared are such that the 'actual' world in the same in both

	LIT (previously 'literal' vs. 'false') 'literal' vs. 'false-kn'	**PSI** (previously 'literal' vs. 'primary') 'literal' vs. 'primary-ign-ambig'	**SSI** (previously 'primary' vs. 'secondary') 'primary-ign-true' vs. 'secondary' 'primary-ign-true' vs. 'secondary-ign' for *some* and *plural*; 'primary-ign-ambig' vs. 'secondary-kn' for *two* and *almost*
SOME	*FTC with full specification* *$\chi^2(1) = 46.8, p = 7.8e\text{-}12$****	*$\chi^2(1) = 18.2, p = 1.9e\text{-}05$****	*FTC with full specification* *$\chi^2(1) = 20.9, p = 4.8e\text{-}06$****
ALMOST	*$\chi^2(1) = 12.0, p = 0.00052$****	*FTC with full specification* *$\chi^2(1) = 20.2, p = 6.9e\text{-}06$****	*$\chi^2(1) = 41.0, p = 1.5e\text{-}10$****
TWO	*$\chi^2(1) = 1.93, p = 0.16$*	*FTC with full specification* *$\chi^2(1) = 6.06, p = 0.014$**	*$\chi^2(1) = 49.2, p = 2.2e\text{-}12$****
PLURAL	*$\chi^2(1) = 23.8, p = .06e\text{-}06$****	*$\chi^2(1) = 0.48, p = 0.49$*	*FTC with full specification* *$\chi^2(1) = 34.3, p = 4.7e\text{-}09$****

Figure 9. Results by subject for the different scalar items



Discussion

This control experiment showed that the presence of a 'secondary-ign' condition did not influence the results: there was no significant difference between the rate of correct answers in 'secondary-kn' and 'secondary-ign'. Second, we still detect the SSI (53% of *no* in condition 'primary'), and if anything at a higher rate than in Experiment 1 (45%). This second experiment thus successfully replicates Experiment 1. Here again, the results cannot be driven by a mere effect of the real-world: there were barely any significant difference across all conditions in the two experiments when comparing sub-conditions in which the real world changes, but not what the speaker knows. The results were stable when ran over the worst sub-condition only, that is, when the information provided in the real-world would go against some implicature.

General discussion

All major theories of implicatures can account for a three-way distinction between the literal reading, the primary reading and the secondary reading. Yet, the experimental literature has not focused on the distinction between the primary and secondary readings, even when it targeted the role of the speaker's information state.

In the following, we first discuss the consequences of our results regarding the debate between Gricean and grammatical accounts: both family of accounts predict that the readings should exist, but they predict that they should arise in different ways. Second, we turn to the comparison of the different items we tested: for *some* and for *almost*, our results suggest that the three readings are available and distinguishable, but numerals and plural pattern differently. Finally, we address a potential objection to the interpretation of our results for *some* and *almost*, and discuss whether they could be explained if only two readings existed.

Role of the Competence Assumption, and the debate between the 'pragmatic' and the grammatical approach.
In the cases that allow to specifically detect the SSI reading ('primary' condition), the speaker is presented as not being *opinionated* about the stronger alternative. As explained in the introduction, on standard pragmatic accounts, the derivation of the SSI crucially depends on the *Competence Assumption:* the assumption that the speaker has an opinion on whether the stronger alternative is true or not. When this assumption is not met, as is the case in the 'primary' condition, listeners should have no ground to reject the sentence: they should not derive the SSI reading. Importantly, we controlled for the fact that participants may judge the sentence according to the actual situation and not the beliefs of the speaker: even in cases where the strong reading of the sentence was true according to the other player's cards ('primary-ign-true' condition), participants rejected the sentence, which shows that they really accessed the SSI. The fact that a sizable proportion of participants provided a *no*-answer in the 'primary' condition can thus be taken to argue against the standard 'pragmatic' approach to SIs, in which the derivation of the SSI is dependent on the opinionatedness of the speaker (the *Competence Assumption*), and would favor the grammatical approach.

These results can be reconciled with a 'pragmatic' approach to SI, however: in recent game-theoretic approaches to SIs, such as the Rational Speech Act model approach (Goodman & Stuhlmüller, 2013; Bergen et al., 2016), the choice of message a speaker makes does not depend just on her information state, but also on her beliefs about the listener's own beliefs about them. Imagine that in the 'primary' condition, where the speaker does not see all the cards, the listener wrongly believes that the speaker sees all the cards. Then it would be inappropriate for the speaker to use *some*, since the listener, believing that the speaker is fully informed, would generate the secondary implicature. Our design leaves underspecified the listener's beliefs about the speaker's belief: there is uncertainty as to what the speaker believes regarding those listener's beliefs. So the rejection of the sentence in the 'primary' condition might find the following explanation: even when the speaker is partially ignorant, she cannot exclude that the listener believes that she is knowledgeable (i.e., that she sees all the cards). In that case, she should not use *some*, because the listener could infer *some but not all*. As pointed out by an anonymous reviewer, alternative accounts that model the role of such higher-order reasoning, such as the RSA account, can thus predict that the *some* utterance would not be chosen (at least, not with high probability) by a speaker who is ignorant about the *all*-alternative - because of the possibility that the listener believes that she is knowledgeable [15].

A design where the speaker's beliefs about the information state of the listener are more controlled or directly manipulated seems hard to implement in a conclusive manner. When the speaker is presented as communicating with a listener who sees that the speaker does not see all the cards ('primary' condition), we would have to ensure that our subjects assume that the speaker knows this. The problem is to rule out the possibility that the subjects who provide a *no*-answer in the 'primary' condition are simply those who are considering the possibility that the speaker is not aware that the listener knows that she is not fully informed.

[15] One might then wonder why we do not get as many *no*-answers in the 'primary' condition (where the speaker is ignorant on whether the sentence with *all* is true) as in the 'literal' condition (where the speaker *knows* that the sentence with *all* is true). RSA-type models are able to distinguish between these two conditions. In the 'primary' condition, the speaker reasons that the listener might draw the wrong conclusion if the listener wrongly believes that the speaker is well informed, but, since the speaker is uncertain about the listener's beliefs, she also assigns a non-zero probability to the possibility that the listener knows that she's seeing only four cards, in which case the listener would not derive the SSI-reading. In contrast, in the 'literal' condition, the speaker knows that the listener is going to derive a false inference whatever his beliefs about the speaker's information state (either that the speaker does not know whether the sentence with `all' holds, or that she believes it to be false).

While we could somehow test explicitly the role played by the beliefs of a fictitious listener about the fictitious speaker's beliefs, by presenting the listener as having more or less visual access to what the speaker sees, it seems much harder to do the same with the higher-order beliefs of the fictitious speaker, i.e. the speaker's beliefs about the listener's beliefs about the speaker's beliefs, which is what matters here.

To conclude, our results suggest that secondary readings can be accessed even when the *Competence Assumption* does not hold. This is not compatible with traditional pragmatic neo-Gricean accounts, and licenses the following disjunctive conclusion: either participants in our task compute the secondary implicature without taking into account speaker's knowledge, as is possible in grammatical approaches, or they engage in sophisticated reasoning about the higher-order beliefs of the speaker, as is possible in pragmatic approaches that take into account higher-order reasoning about the speaker's beliefs about the listener's beliefs about the speaker's beliefs about the world. We could then conclude that either secondary readings can be derived in the grammar, as part of the compositional semantics, or their derivation involves the kind of sophisticated higher-order reasoning that is posited in the Rational Speech Act model and related frameworks.

Comparison of items

For *some* and *almost*, our paradigm enables to distinguish three distinct readings. For *almost*, this is a new result: the existence of a literal reading ("Almost all cards are hearts" being accepted when all cards are hearts) had to our knowledge never been established. (Note that we tested two cases for *almost*, *almost all* and *almost no*: the rate of strengthened meanings is a little higher with *almost all* than with *almost no*, but the same pattern shows up with both). Our results thus support the SI account of *almost* (Sadock, 1981; Spector, 2012), and challenge the accounts of the 'not all' meaning component as a logical entailment (Hitzeman, 1992; Horn, 2011; Kilbourn-Ceron, 2015). This conclusion is further supported by the striking similarities of behavior between *some* and *almost*: even if the rates of derivation differ (20% of LIT for *almost*, 35% for *some*) - which is unsurprising given the variability across scalar items (van Tiel et al., 2013), we find the same distribution of PSI and SSI overall.

Regarding numerals and plural, we find a different pattern of answers, that suggest that if PSI exists at all for these items, it is at least less accessible than with *some* and *almost* (cf. Tables 5 and 8 which report tests of the relevant interactions). For numerals, this adds to other differences already found with "standard" scalar items, regarding syntactic distribution (Horn, 1992; Breheny, 2008), acquisition (Papafragou & Musolino, 2003; Huang & Snedeker, 2009) and processing (Huang & Snedeker, 2009; Marty et al., 2013), and challenges the traditional SI account proposed by Horn (1972). Our results suggest in a similar way that the 'strictly more than one' meaning component of the plural morpheme does not involve the same mechanisms as the strong readings of "standard" scalar items. The parallel with numerals remains to be explored. In both cases, our results are consistent with the view that the two readings stem from an ambiguity at the level of literal meanings, as has been suggested in the literature (see Breheny, 2008; Kennedy, 2015 for numerals, and Farkas & Swart, 2010; Križ, 2017 for plural[16]), rather than an account in terms of SIs. Note however that, in a priming study, Bott & Chemla (2016) found that the strong readings of numerals and *some* can 'prime' each other, but not the strong reading of plural expressions, suggesting a different typology whereby the strong reading of numerals and *some* have more in common with each other than with that of *plural.* This study was not designed to test how these items compare to each other regarding primary implicatures: it might be that the mechanism giving rise to strong readings are at some level analogous for numerals and for *some* (involving for instance the consideration of focal alternatives), and yet different at some other level. (For instance, the *literal* meaning of numerals might include a maximality component which has some formal similarities with exhaustification (Kennedy, 2015), in which the semantics of numerals include a maximality component where maximality is relative to the set of degrees numerals quantifies over). One possibility is that the strong reading of numerals is a scalar implicature, but that the difference with *some* is that numerals are intrinsically focused and give rise to the strong, exhaustified meaning by default, even in environments that usually do not support them (see Spector, 2013). Moreover, in a recent study building on ours, Cremers et al. (2017) found evidence for primary implicatures for bare numerals. Their results might suggest that primary implicatures are not impossible with numerals, but less accessible than with other items, maybe due to the salience of strong readings.

A potential objection: could we explain the results with just two readings?

A reviewer pointed out that even in the cases of *some* and *almost*, one could try to explain our results with only two readings, the literal reading and the SSI reading, and no PSI reading. In a sense, this is a higher-order version of the objection we raised previously against studies which only distinguished between two readings. We will explain the objection and present two arguments to mitigate it. In short, the objection relies on an interpretation of the task that does not seem particularly likely to us and, empirically, the absence of the three-

[16] Cf. footnote 5.

way distinction for numeral and plural suggests that in cases where this distinction occurs, it is indeed due to implicatures and not to extrinsic facts about the task, which would be expected to apply across the board.

Assume that there are only two readings: the literal reading and the SSI reading. Some participants could reason in the following manner: in the 'primary' condition, the speaker might believe that the listener knows that she is not opinionated about the *all* alternative (as the listener might see that the speaker does not see all the cards). In that case, the listener (as perceived by the speaker), knowing that the speaker is not opinionated, would not derive a reading that would entail that she (the speaker) is, and so should go for the literal meaning. For the speaker, there would be no risk that the *some* sentence gives rise to the SSI reading, so she should be able to use the sentence. Participants who reason in this way would then answer *yes* in the 'primary' condition. However, other participants might make other assumptions about the higher-order beliefs of the speaker: for instance, that the speaker does not exclude that the listener believes that she is opinionated about the *all* alternative. In this case, they would answer *no* if they give a significant weight to the SSI reading. This would explain the intermediate rating of the 'primary' condition, without any need to posit three distinct readings.

This interpretation of our results does not seem very plausible to us. For it to work, the 'primary' condition has to be construed as one where the speaker might believe that the listener knows (or might know) that the speaker is not opinionated about the *all* alternative. However, even if we imagine that the listener observes that the speaker does not see all the cards, the listener is not thereby able to exclude the possibility that the speaker believes that *some but not all cards are hearts*: the speaker might well see (for all the listener knows), say, three hearts and one spade (in which case the SSI reading of *Some of the cards are hearts* would be true). The important point is that even in the so-called *Ignorance* condition, the speaker *could* be opinionated about the *all* alternative. In fact, this type of situation is instantiated by the additional control 'secondary' condition ('secondary-ign') added in our second experiment, and also by some of the cases instantiating conditions 'literal' and 'primary' with numerals and some of the conditions associated with sentences with FIRST and LAST in our second experiment. If we assume that, in all conditions, the speaker believes that the listener sees how many cards the speaker is seeing, and does not have any other source of knowledge about her information state, then in the 'primary' condition the speaker should not be more prone to use the *some* sentence than in the 'literal' condition: in both cases, if only the literal reading and the SSI reading exist, the listener might conclude 'some but not all' (by picking the SSI reading), since the listener cannot exclude that the speaker believes 'some but not all'.

There might be diverse dependencies between the first-order properties of the experimental conditions (which cards are seen by the speaker) and the participants' hypotheses regarding the higher-order beliefs of the speaker about the listener. But the particular dependency needed here does not seem particularly well motivated. Some participants but not others would need to conceptualize the 'primary' condition as one where the listener might know that the speaker is not opinionated about the stronger alternative (incorrectly if we imagine that the listener just knows how many cards the speaker sees), despite the presence of similar conditions where the speaker is in fact opinionated (e.g. the 'secondary-ign' conditions in Experiment 2, where the speaker also sees only half of the cards), while all participants would have to conceptualize the 'literal' and 'secondary' conditions in the same way, as one where the listener knows that the speaker is opinionated.

Furthermore, and maybe more directly, the comparison of the various items provides evidence against the alternative interpretation in terms of a two-way ambiguity for *some* and *almost*. Under the proposed explanation, there should be no difference between numerals and plural, on the one hand, and *some* and *almost*, on the other hand. In the case of numerals and plural, we did not find any difference between the 'literal' and 'primary' conditions. This is readily explained if numerals and plural, in contrast with *some*, only license the literal and SSI-reading, or (in the case of numerals) make the PSI reading much less accessible. Under the proposed explanation, we should get an intermediate rating for the 'primary' condition no matter what, and to the same extent as with *some*. In both cases, there would be only two readings, and we would expect to see the same dependencies between first-order facts about the speaker's knowledge of the cards and higher-order facts about the speaker's beliefs about the listener's beliefs.

General summary

Most experimental studies of SI are based on a distinction between the literal meaning (LIT) and interpretations with a scalar implicature (SI). Our method allows us to tease apart the three readings discussed in the theoretical literature, by distinguishing between primary (PSI) and secondary (SSI) implicatures. For the "standard" case of scalar implicatures, *some*, we find, as expected, evidence for three distinct interpretations, a result replicated across 3 experiments.[17] Our results suggests that secondary implicatures are accessible even in contexts where the speaker is presented as lacking information, a result that is not expected on classical neo-Gricean pragmatic

[17] A third experiment involving a dual-task paradigm is presented in appendices (Appendix C).

accounts of SI. These findings are compatible with both the grammatical approach to SIs and more recent game-theoretic models of pragmatics, where participants engage in sophisticated higher-order reasoning about each other. Another goal of the study was to compare various items whose interpretation have been argued to involve scalar implicatures: *almost*, plural morphology and numerals. We found that the negative implication associated with *almost* behaves in the same way as the one associated to *some*, a finding that supports the SI approach. For the other two cases, numerals and plural morphology, we found no evidence for PSI, which adds to previous findings suggesting that these cases depart from standard cases.

Acknowledgments
The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013) / ERC Grant Agreement n.313610 and was supported by ANR-10-IDEX-0001-02 PSL* and ANR-10-LABX-0087 IEC.

References

Baddeley, A. (1992). Working memory. *Science*, *255*(5044), 556-559.

Bale, A., & Barner, D. (2013). Grammatical alternatives and pragmatic development. Alternatives in Semantics,'Studies in Pragmatics, Language and Cognition'. Palgrave Macmillan. New York, 238-66.

Barner, D. & Bachrach, A. (2010). Inference and exact numerical representation in early language development. *Cognitive Psychology*, 60, 40–62.

Bates D, Maechler M, Bolker B and Walker S (2014). _lme4: Linear mixed-effects models using Eigen and S4_. R package version 1.1-7, < URL: http://CRAN.R-project.org/package = lme4>.

Bates D, Maechler M, Bolker BM and Walker S (2014). "lme4: Linear mixed-effects models using Eigen and S4." ArXiv e-print; submitted to _Journal of Statistical Software_, < URL: http://arxiv.org/abs/1406.5823>.

Bergen, L., & Grodner, D. J. (2012). Speaker knowledge influences the comprehension of pragmatic inferences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 38, 1450-1460.

Bergen, L., R. Levy, & Goodman, N. (2016). Pragmatic reasoning through semantic inference. *Semantics and Pragmatics* 9.

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.

Bott, L., Bailey, T. M., & Grodner, D. (2012). Distinguishing speed from accuracy in scalar implicatures. *Journal of Memory and Language*, *66*(1), 123-142.

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.

Breheny, Richard. (2008) "A new look at the semantics and pragmatics of numerically quantified noun phrases." *Journal of Semantics* 25.2 : 93-139.

Buccola, B. & Haida, A. (in press), "Obligatory irrelevance and the computation of ignorance inferences", to appear in the *Journal of Semantics*.

Chemla, E., & Bott, L. (2015). Using structural priming to study scopal representations and operations. *Linguistic Inquiry*.

Chemla, E., & Singh, R. (2014). Remarks on the experimental turn in the study of scalar implicature, Part I. *Language and Linguistics Compass*, *8*(9), 373-386.

Chierchia, G. (2004). Scalar implicatures, polarity phenomena, and the syntax/pragmatics interface. *Structures and beyond*, *3*, 39-103.

Chierchia, G. (2006). Broaden your views: Implicatures of domain widening and the "logicality" of language. *Linguistic inquiry*, *37*(4), 535-590.

Chierchia, G., Fox, D., & Spector, B. (2009). Hurford's constraint and the theory of scalar implicatures. *Presuppositions and implicatures*, *60*, 47-62.

Chierchia, G., Fox, D., & Spector, B. (2012). Scalar implicature as a grammatical phenomenon. In . P. Portner, C. Maienborn et K. von Heusinger (eds.) *An International Handbook of Natural Language Meaning Semantics*, vol 3: 2297-2331, Mouton de Gruyter

Cremers, A., Coppock, ., Dotlacil, J, & Roelofsen, F. (2017). Modified numerals: two routes to ignorance. Ms. https://semanticsarchive.net/Archive/jk2M2Q3O/paper.pdf.

De Neys, W. (2006). Dual processing in reasoning two systems but one reasoner. *psychological Science*, *17*(5), 428-433.

De Neys, W., & Schaeken, W. (2007). When people are more logical under cognitive load. *Experimental Psychology (formerly Zeitschrift für Experimentelle Psychologie)*, *54*(2), 128-133.

Degen, J., & Tanenhaus, M. K. (2011). Making inferences: the case of scalar implicature processing. In *Proceedings of the 33rd annual conference of the cognitive science society* (pp. 3299-3304). Austin, TX: Cognitive Science Society.

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.
Engle, R. W. (2002). Working memory capacity as executive attention. *Current directions in psychological science*, *11*(1), 19-23.
Farkas, D. & H. de Swart (2010). The semantics and pragmatics of plurals. *Semantics and Pragmatics* 3.
Form, L. (1995). Plurality, conjunction and events.
Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.
Fox, D. (2007). Free choice and the theory of scalar implicatures. In Uli Sauerland and Penka Stateva (Eds.), *Presupposition and implicature in compositional semantics*. Houndmills, Basingstoke: Palgrave Macmillan.
Fox, D. (2014). Cancelling the Maxim of Quantity: Another challenge for a Gricean theory of scalar implicatures. *Semantics and Pragmatics*, *7*, 5-1.
Franke, Michael. "Quantity implicatures, exhaustive interpretation, and rational conversation." Semantics and Pragmatics 4 (2011): 1-1.
Gazdar, G. (1979). Pragmatics: Implicature, presupposition, and logical form. New York: Academic Press.
Geurts, B. (2010). *Quantity implicatures*. Cambridge University Press.
Goodman, N. D. & A. Stuhlmüller (2013). Knowledge and implicature: Modeling language understanding as social cognition. *Topics in cognitive science 5* (1), 173–184.
Grice, P. (1975). Logic and conversation. In P. Cole & J. Morgan (Eds.), *Syntax and Semantics,* Volume 3. New York: Academic 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.
Hamblin, Charles L. "Questions in montague english." *Foundations of language* 10.1 (1973): 41-53.
Hirschberg, J. L. B. (1985). *A theory of scalar implicature*. University of Pennsylvania.
Hitzeman, J. (1992). The selectional properties and entailments of almost. In*Papers from the 28 th Regional Meeting of the Chicago Linguistic Society* (pp. 225-238)..
Hochstein, L.; Bale, A.; and Barner, D. (2017) Scalar implicature in absence of epistemic reasoning? The case of Autism Spectrum Disorders. Language Learning and Development, DOI: 10.1080/15475441.2017.1343670.
Horn, L. (1972). On the Semantic Properties of Logical Operators in English. Ph.D. dissertation. University of California. Los Angeles, CA.
Horn, L. R. (2011). Almost forever. Pragmatics and autolexical grammar: In honor of Jerry Sadock, 3-21.
Huang, Y. T., & Snedeker, J. (2009). Online interpretation of scalar quantifiers: Insight into the semantics–pragmatics interface. *Cognitive psychology*, *58*(3), 376-415.
Huang, Y. T., & Snedeker, J. (2009). Semantic meaning and pragmatic interpretation in 5-year-olds: evidence from real-time spoken language comprehension. *Developmental psychology*, *45*(6), 1723.
Ivlieva, N. (2013). Scalar implicatures and the grammar of plurality and disjunction. PhD dissertation, MIT.
Katzir, Roni. "Structurally-defined alternatives." *Linguistics and Philosophy*30.6 (2007): 669-690.
Kennedy, C. (2015). A "de-Fregean" semantics (and neo-Gricean pragmatics) for modified and unmodified numerals. *Semantics and Pragmatics* 8.
Kilbourn-Ceron, O. (2015). Embedded exhaustification: evidence from almost.
Križ, M. (2017). Bare Plurals, Multiplicity and Homogeneity. Ms. Institut Jean Nicod.
Landman, F. (1998). *Plurals and maximalization* (pp. 237-271). Springer Netherlands.
Lasersohn, Peter. "Plurality." (1995).
Levinson, S. C. (2000). Presumptive meanings: The theory of generalized conversational implicature. Cambridge, MA: MIT press.
Marty, P., & Chemla, E. (2013). Scalar implicatures: working memory and a comparison with only. *Frontiers in psychology*, *4*.
Marty, P., Chemla, E., & Spector, B. (2013). Interpreting numerals and scalar items under memory load. *Lingua*, *133*, 152-163.
Meyer, M-C. (2013). *Ignorance and Grammar*. PhD dissertation, MIT.
Meyer, M-C. (2014). Deriving Hurford's Constraint. *Proceedings of SALT* 24: 577-596. doi: 10.3765/salt.v24i0.2518
Miyake, A., & Shah, P. (Eds.). (1999). Models of working memory: Mechanisms of active maintenance and executive control. Cambridge University Press.
Noveck, I. (2001), When children are more logical than adults: Experimental investigations of scalar implicature. *Cognition,* 78, 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: 253-282.

Pearson, H. A., Khan, M., & Snedeker, J. (2011, March). Even more evidence for the emptiness of plurality: An experimental investigation of plural interpretation as a species of implicature. In *Proceedings of SALT* (Vol. 20, pp. 489-507).
R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.
R Core Team (2013). R: A language and environment for statistical computing. R
Rips, L. J. (1975). Inductive judgments about natural categories. *Journal of verbal learning and verbal behavior*, *14*(6), 665-681.
Russell, B. (2006). Against grammatical computation of scalar implicatures.*Journal of semantics*, *23*(4), 361-382.
Russell, B. (2012). Probabilistic reasoning and the computation of scalar implicatures. PhD dissertation, Brown University.
Sadock, J. (1981). Almost. *Radical pragmatics*, 257-271.
Sauerland, U. (2004). Scalar Implicatures in Complex Sentences. *Linguistics and Philosophy,* 27, 367-391
Sauerland, U., Anderssen, J., & Yatsushiro, K. (2005). The plural is semantically unmarked. *Linguistic evidence*, 413-434.
Sauerland, U. (2012). The computation of scalar implicatures: Pragmatic, lexical or grammatical?. Language and Linguistics Compass, 6(1), 36-49.
Smith, C. L. (1980). Quantifiers and question answering in young children.*Journal of Experimental Child Psychology*, *30*(2), 191-205.
Spector, B. (2003). Scalar Implicatures: Exhaustivity and Gricean Reasoning. *Proceedings of the ESSLLI*, 3, 277-288.
Spector, B. (2007). Aspects of the pragmatics of plural morphology: On higher-order implicatures. *Presuppositions and implicatures in compositional semantics*, 243-281.
Spector, B. (2013). Bare numerals and scalar implicatures. Language and Linguistics Compass, 7(5), 273-294.
Spector, B. (2014). Global positive polarity items and obligatory exhaustivity.*Semantics and Pragmatics*, *7*(11), 1-61.
Tieu, L., Bill, C., Romoli, J., & Crain, S. (2014, September). Plurality inferences are scalar implicatures: Evidence from acquisition. In *Semantics and Linguistic Theory* (Vol. 24, pp. 122-136).
Tomlinson, J., Bott, L., & Bailey, T. (2011). Understanding literal meanings before pragmatic inference: Mouse-trajectories of scalar implicatures. In *th Biennial Conference of Experimental Pragmatics, June* (pp. 2-4).
Van Rooij, R., & Schulz, K. (2004). Exhaustive interpretation of complex sentences. *Journal of logic, language and information*, *13*(4), 491-519.
Van Tiel, B., Van Miltenburg, E., Zevakhina, N., & Geurts, B. (2014). Scalar diversity. *Journal of Semantics*, ffu017.
Zweig, E. (2008). Dependent plurals and plural meaning. ProQuest.

Appendices

Appendix A. Displayed cards depending on the condition
Mary: "X cards are [hearts]".
Kn: the speaker is knowledgeable.
Ign: the speaker is ignorant (in the sense of not seeing all the cards).

	Target conditions						Control conditions									
Conditions	literal		primary-ign-true		primary-ign-ambig		false-kn		false-ign-true		false-ign-false		secondary-kn		secondary-ign	
Knowledge state of the speaker	Kn		Ign		Ign		Kn		Ign		Ign		Kn		Ign	
Actual world	if LIT: 'true' if SI; 'false'		actually true		if LIT: 'true' if SI; 'false'		actually false		actually true		actually false		actually true		actually true	
Player	Mary	Peter	Mary	Peter	Mary	Peter	Mary	Peter	Mary	Peter	Mary	Peter	Mary	Peter	Mary	Peter
SOME	♥♥ ♥♥ ♥♥ ♥♥	♥♥ ♥♥ ?? ??	♥♥ ♥♥ ?? ??	♥♥ ♥♥ ♤♤♤ ♤	♥♥ ♥♥ ?? ??	♥♥ ♥♥ ♥♥ ♥♥	♤♤♤ ♤ ♤♤♤ ♤	♤♤♤ ♤ ?? ??	♤♤♤ ♤ ?? ??	♤♤♤ ♤ ♥♥ ♥♥	♤♤♤ ♤ ?? ??	♤♤♤ ♤ ♤♤♤ ♤	♥♥ ♥♥ ♤♤♤ ♤	♥♥ ♥♥ ?? ??	♥♥ ♥♤ ?? ?♤	♥♥ ♥♤ ♥♥ ♥♤
PLURAL	♥♤ ♤♤ ♤♤♤ ♤	♥♤ ♤♤ ?? ??	♥♤ ♤♤ ?? ??	♥♤ ♤♤ ♥♥ ♥♤	♥♤ ♤♤ ?? ??	♥♤ ♤♤ ♤♤♤ ♤	♤♤♤ ♤ ♤♤♤ ♤	♤♤♤ ♤ ?? ??	♤♤♤ ♤ ?? ??	♤♤♤ ♤ ♥♥ ♥♥	♤♤♤ ♤ ?? ??	♤♤♤ ♤ ♤♤♤ ♤	♥♥ ♥♥ ♤♤♤ ♤	♥♥ ♥♥ ?? ??	♥♥ ♥♤ ??? ♤	♥♥ ♥♤ ♥♥ ♥♤
TWO	♥♥ ♥♥ ♤♤♤ ♤	♥♥ ♥♥ ?? ??	♥♥ ♤♤ ?? ??	♥♥ ♤♤ ♤♤♤ ♤	♥♥ ♤♤ ?? ??	♥♥ ♤♤ ♥♥ ♤♤	♤♤♤ ♤ ♤♤♤ ♤	♤♤♤ ♤ ?? ??	♥♤ ♤♤ ?? ??	♥♤ ♤♤ ♥♤ ♤♤	♥♤ ♤♤ ?? ??	♥♤ ♤♤ ♥♤ ♤♤	♥♥ ♤♤ ♤♤♤ ♤	♥♥ ♤♤ ?? ??	-	-
ALMOST ALL	♥♥ ♥♥ ♥♥ ♥♥	♥♥ ♥♥ ?? ??	♥♥ ♥♥ ♥♥ ♥?	♥♥ ♥♥ ♥♥ ♥♤	♥♥ ♥♥ ♥♥ ♥?	♥♥ ♥♥ ♥♥ ♥♥	♤♤♤ ♤ ♤♤♤ ♤	♤♤♤ ♤ ?? ??	♤? ?? ?? ??	♤♥ ♥♥ ♥♥ ♥♥	♤? ?? ?? ??	♤♤♤ ♤ ♤♤♤ ♤	♥♥ ♥♥ ♥♥ ♥♤	♥♥ ♥♥ ?? ??	-	-
ALMOST NO	♤♤♤ ♤ ♤♤♤ ♤	♤♤♤ ♤ ?? ??	♤♤♤ ♤ ♤♤♤ ?	♤♤♤ ♤ ♤♤♤ ♥	♤♤♤ ♤ ♤♤♤ ?	♤♤♤ ♤ ♤♤♤ ♤	♥♥ ♥♥ ♥♥ ♥♥	♥♥ ♥♥ ?? ??	♥? ?? ?? ??	♥♤ ♤♤ ♤♤♤ ♤	♥?? ? ?? ??	♥♥ ♥♥ ♥♥ ♥♥	♥♤ ♤♤ ♤♤♤ ♤	♥♤ ♤♤ ?? ??	-	-
ALL	-	-	-	-	-	-	♥♥ ♥♥ ♤♤♤ ♤	♥♥ ♥♥ ?? ??	♥♥ ♥♥ ?? ??	♥♥ ♥♥ ♥♥ ♥♥	♥♥ ♥♥ ?? ??	♥♥ ♥♥ ♤♤♤ ♤	♥♥ ♥♥ ♥♥ ♥♥	♥♥ ♥♥ ?? ??	-	-
NO	-	-	-	-	-	-	♥♥ ♥♥ ♤♤♤ ♤	♥♥ ♥♥ ?? ??	♤♤♤ ♤ ?? ??	♤♤♤ ♤ ♤♤♤ ♤	♤♤♤ ♤ ?? ??	♤♤♤ ♤ ♥♥ ♥♥	♤♤♤ ♤ ♤♤♤ ♤	♤♤♤ ♤ ?? ??	-	-
FIRST	-	-	-	-	-	-	♤♤ ♥♥ ♤♤ ♥♥	?? ♥♥ ?? ♥♥	-	-	?? ♥♥ ?? ♥♥	♤♤ ♥♥ ♤♤ ♥♥	♥♥ ♤♤ ♥♥ ♤♤	♥♥ ♤♤ ?? ??	♥♥ ♤♤ ?? ??	♥♥ ♤♤ ♥♥ ♤♤
LAST	-	-	-	-	-	-	♥♥ ♤♤ ♥♥ ♤♤	♥♥ ?? ♥♥ ??	-	-	♥♥ ?? ♥♥ ??	♥♥ ♤♤ ♥♥ ♤♤	♤♤ ♥♥ ♤♤ ♥♥	?? ?? ♤♤ ♥♥	?? ?? ♤♤ ♥♥	♤♤ ♥♥ ♤♤ ♥♥

<u>Remark</u>: 'primary' always corresponds to 'literal' with some cards put face down.

Appendix B. Number of trials by condition

		Target conditions			Control conditions					Total		
Condition		literal	primary-ign-true	primary-ign-ambig	false-kn	false-ign-true	false-ign-false	secondary-kn	secondary-ign	ex	trials	all
Experiment 1	SOME	8	4	4	8	4	4	16	-	18	288	306
	PLURAL	8	4	4	8	4	4	16	-			
	TWO	8	4	4	8	4	4	16	-			
	ALMOST ALL	8	4	4	8	4	4	16	-			
	ALMOST NO	8	4	4	8	4	4	16	-			
	ALL	-	-	-	4	4	4	12	-			
	NO	-	-	-	4	4	4	12	-			
Experiment 2	SOME	8	4	4	8	4	4	8	8	18	320	338
	PLURAL	8	4	4	8	4	4	8	8			
	TWO	8	4	4	8	4	4	16	-			
	ALMOST ALL	8	4	4	8	4	4	16	-			
	ALMOST NO	8	4	4	8	4	4	16	-			
	ALL	-	-	-	4	4	4	12	-			
	NO	-	-	-	4	4	4	12	-			
	FIRST	-	-	-	4	-	4	4	4			
	LAST	-	-	-	4	-	4	4	4			
Dual task experiment*	SOME	8	4	4	4	2	2	4	4	4+4	184	192
	PLURAL	8	4	4	4	2	2	4	4			
	TWO	8	4	4	4	2	2	8	-			
	ALMOST ALL	8	4	4	4	2	2	8	-			
	ALMOST NO	8	4	4	4	2	2	8	-			
	ALL	-	-	-	2	2	2	6	-			
	NO	-	-	-	2	2	2	6	-			

* to be divided by 2, according to the two Cognitive Load conditions

Appendix C. Dual-task experiment.

Goal

Diverse methodologies have been used to study the processing properties of SI, most of them showing that there is a cost associated to SI computation (Bott & Noveck, 2004; De Neys & Schaeken, 2007; Posada & Noveck, 2003; Rips, 1975, a.o.). For example, the study of response times in classical truth value judgment tasks (Bott & Noveck, 2004; Posada & Noveck, 2003, a.o.) showed that in target sentences, participants who judged the statements to be false were slower than those who judged them to be true. In the same way, the rate of SI increased as a permitted response time did. That suggests that SI is derived *from* LIT, in a sequential way, a result highly coherent under the Neo-Gricean approach but also compatible with a Grammatical approach. This finding was confirmed using other methodologies: dual-task paradigm, eye tracking studies, mouse tracking studies: SI are derived with a delay (Bott, Bailey, & Grodner, 2012).

In order to test the processing properties of the three interpretations presented in the introduction, a dual task version of the experiment was implemented. Its main goal was to study the processing properties of the three levels of reading established in the first experiments, and orthogonally to have a better understanding of the cost traditionally associated to SI. As in the first experiments, we compared the behavior of four scalar items: SOME, ALMOST, NUMERALS and PLURAL.

Theoretical background

The reasoning underlying a dual-task experiment is based on the working memory model, which will not be presented in detail here (but see Baddeley, 1992; Miyake & Shah 1999; Engle, 2002). Crucially, the reasoning relies on the assumption that human executive cognitive resources (working memory resources) are limited: introducing a second task reduces the resources available for the first task, facilitating automatic responses and inhibiting analytic responses (de Neys, 2006). Going back to SI, this offers a way to reveal the 'basic' meaning of a sentence by blocking potential strengthening mechanisms. From a methodological point of view, the paradigm is based on the comparison between two conditions of Cognitive Load, a factor characterizing the degree to which working memory resources are burdened: a LOW LOAD condition (with an easy second task), a HIGH LOAD condition (with a harder second task). We also added a baseline condition without the second task (NO LOAD).

To our knowledge, four studies have tested SI using a dual-task paradigm (De Neys & Schaeken, 2007, Dieuassert et al., 2011; Marty & Chemla, 2011; Marty et al., 2013). They have shown that, for sentences containing SOME, participants derive less SI as the second task becomes harder. In line with other studies using response time, self-paced reading or visual-world paradigm, it confirms that SI are costly as compared to LIT, a result that was first used to argue against proposals suggesting that SI could be generated automatically and then optionally cancelled as need be (e.g. Levinson, 2000). It further enables to bring more precise information on the nature of the cost, indicating whether the cognitive effort associated with the processing involves central working memory resources (whereas the conclusion we can draw from Response Time studies is that SI are derived *later* than LIT and does not characterize the nature of the resources involved). In the four studies mentioned above, PSI and SSI are confounded: the dual-task could indicate at which level of SI processing working memory resources were specifically involved.

The second goal was to compare the behavior of four scalar items. Interestingly, Marty et al. (2013) had found that the effect of the dual task was reversed for NUMERALS and SOME: participants accessed more the 'exactly n' reading (assumed implicature) under high cognitive load, whereas for SOME, they accessed *less* the implicature ('some but not all'). The paradigm used was almost exactly the same as Marty et al.'s.

Method and materials

The experiment was an online experiment, hosted on Alex Drummond's Ibex Farm. Participants were recruited via Mechanical Turk and were paid for their participation. Participants had to do two tasks at the same time:

- a truth value judgment task, identical to the task implemented in the first experiment.
- a letter memory task, very similar to the task implemented by Marty et al. (2013).

Truth Value Judgment Task

The task was identical to the task implemented in the first experiments and presented before.

Letter Memory Task

The memory task was a short term storage task of sequences of letters, based on the task implemented by Marty et al. Before the Truth Value Judgment Task, a sequence of letters was shown to the participants. The letters were presented one after the other for 800ms, with 50ms pause between them. They were displayed in the center of the screen, in black, in upper case. The sequences were generated randomly using a program in Python. 9 letters were used: B, F, H, J, L, M, Q, R and X (chosen to be phonologically distinct). After the Truth Value Judgment Task, participants had to give back the sequence of letters in reverse order. They were given feedback at the end of the trial: either “Correct”, displayed in green in the center of the screen, or “Wrong”, displayed in red, with an error message (e.g., “You typed DL and the correct answer was LF.”)

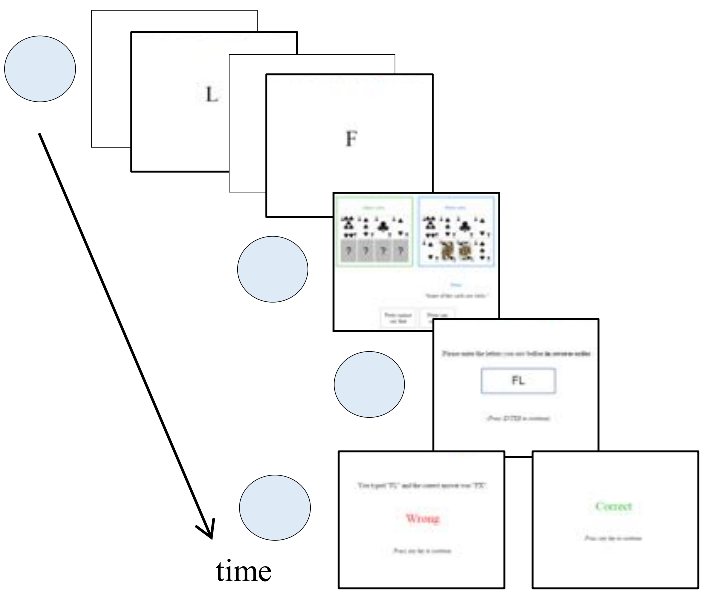
Cognitive Load was manipulated by varying the length of the sequence of letters: 2 letters in the LOW LOAD and 4 letters in the HIGH LOAD condition. Memory resources were supposed to be more heavily taped in the HIGH condition. The cognitive load was manipulated within subjects: each participant performed the LOW LOAD as well as the HIGH LOAD task. A control experiment, with no dual task, was also implemented as a baseline. Participants in the dual-task version of the experiment were administered two blocks of 92 trials, with a short break between them: one block contained LOW LOAD trials, and the other block contained HIGH LOAD trials. The order of the blocks was randomly determined for each participant. In each block, the order of items and the correspondence between the sequence of letters and the truth-value judgment task item was generated randomly.

Due to the difficulty of the dual task, the number of trials from Experiment 2 (320 trials) had to be reduced. Given that previous results indicated that the presence of a ‘secondary-ign’ condition did not strongly influence the judgments, FIRST and LAST control sentences were removed, but there was still the subdivision of ‘secondary’ into a Kn and a Ign condition. In order to reduce the number of trials, half of the ‘false’ and ‘secondary’ conditions (*i.e.* control trials) were removed. The proportion of target cases in the experiment was thus higher than in Experiment 2: there was 25% ‘false’, 25% ‘literal’, 25% ‘primary’, 25% ‘secondary’ (*i.e.* 50% of target cases). There were 184 trials in total.

Procedure

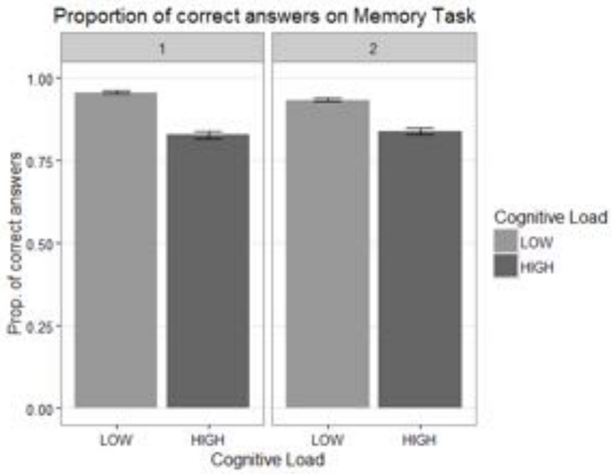
The experiment was an online experiment, hosted on Alex Drummond's Ibex Farm. After having given their consent to participate in the experiment, participants were given instructions concerning the Truth Value Judgment Task only. There was then a first training (4 non ambiguous sentences with feedback). Then, participants were given the second part of the instructions, concerning the Letter Memory Task. There was then a second training phase *with* the Memory Task, on the 4 same sentences as before. The sequences were composed of 2 or 4 letters depending on the block they started with. The experimental phase was divided in two blocks, according to the cognitive load (LOW LOAD vs. HIGH LOAD). Participants were asked to make a short break between them. The two first sentences after the break were taken from the training phase, in order to get the subjects used with the new number of letters to memorize. Each trial started with the presentation of the sequence of letters. Then, the Truth Value Judgment task was displayed. It remained until the participant answered. Next, participants had to reproduce the sequence of letters in reverse order. Last, they were given feedback on the accuracy of their answer (see Figure 11). At the end of the experiment, there was a short questionnaire (information on age, sex, native language, kind of device used to answer, Mechanical Turk Worker ID, and kind of strategy used to memorize the letters). This last question was included in order to control for the fact that participants may have written the letters. For the no dual task version of the experiment, the procedure was exactly the same, except that the instructions and the final questionnaire were adapted to the task.

Figure 11. Description of the dual-task procedure. (1) Presentation of the sequence of letters (2 or 4 letters) ; (2) Truth Value Judgment Task; (3) Letter Memory Task (give back the sequence in reverse order); (4) Feedback on Letter Memory Task.



Participants
59 participants were recruited via Mechanical Turk for the dual-task version of the experiment, and 61 for the no-dual task version of the experiment. In the dual-task version, 4 participants were removed due to a problem loading their data, 4 because they made more than 18% of errors on controls (2 standard deviations from the mean), and 1 who indicated that he had written the letters down to perform the memory task. All participants reported English as their native language. In the no-dual task, we removed from the analysis 3 participants whose native language was not English, 3 because they made more than 20% of errors on controls and 2 because they made more than 20% of errors on NO-controls. Thus, there was 53 participants for the dual task version of the experiment, and 50 participants for the baseline without the dual task (57 females, 46 males, mean age: 39,7, from 19 to 69 year old).
Results
Data analyses were conducted using R. We used binomial linear mixed effects model, built with a maximal random effect structure based on subjects and items as random variables, although we sometimes had to step back to random-intercepts-only models when the model failed to converge with the full random-effects specification (following Barr et al., 2013).
Letter Memory Task
The mean rate of correct answers on the Memory Task was overall quite high (89%). As expected, there was a significant effect of the Cognitive Load condition (see figure 12): participants made more errors on HIGH LOAD trials than on LOW LOAD trials (HIGH LOAD: 83% (SD:.03) vs. LOW LOAD: 94 % (SD:.01) ($\chi^2(1) = 37, p = 9.9e\text{-}10$***) . This confirmed that the 4-letter sequences were more demanding than the 2-letter sequences. There was no effect of the order of the blocks (χ^2 $(1) = 0.48, p = .49$).

Figure 12. Proportion of correct answers on the Memory Task (by Cognitive Load condition and order of blocks)



Tradeoff analysis
We also tested whether there was an effect of the answer on the Truth-Value Task on the Memory Task results. Indeed, the cost associated with the SI computation can show up on the rate of correct answers on the Memory Task. This corresponded to the following hypothesis; in condition 'literal' and 'primary', there should be more errors if participants previously answered *no* (*i.e.* the implicature was derived) than if they previously answered *yes*; in contrast, in control conditions, the previous answer should not affect the rate of errors. The results were not significant whatever the level of Cognitive Load (LOW, HIGH or merged): there was no effect of the answer given on TVJT (condition 'literal', all scalar items confounded, all Cognitive Load Condition: χ^2 $(1) = 0.18, p = 0.67$; condition 'primary': χ^2 $(1) = 0.10, p = 0.75$). There was no effect either when tested by scalar item. This means that there was no tradeoff between the two tasks.

Truth value judgment task
As summed up in Table 10, the results of the first experiments were replicated (except that for TWO the difference between 'literal' and 'primary' is significant). The interaction between TWO and all other cases was now significant and the interaction between SOME and ALMOST was the only non-significant interaction ($\chi^2(1) = 0.69, p = .40$). Table 11 present the results for the dual-task experiment (trials with incorrect answers on the *Memory Task* were removed). It also replicates findings of the first experiments. These results are illustrated in Figure 13 and Figure 14.

Table 10. Three readings for target items (no dual task experiment)

No dual task	**LIT:** 'false' vs. 'literal'	**PSI:** 'literal' vs. 'primary'	**SSI:** 'primary' vs. 'secondary'
All items	$\chi^2(1) = 60, p = 7.9e\text{-}15$ ***	$\chi^2(1) = 33, p = 1.1e\text{-}08$ ***	$\chi^2(1) = 61, p = 6.7e\text{-}15$ ***
SOME	$\chi^2(1) = 43, p = 5.9e\text{-}11$ ***	$\chi^2(1) = 41, p = 1.4e\text{-}10$ ***	$\chi^2(1) = 24, p = 1.1e\text{-}06$ ***
ALMOST	$\chi^2(1) = 7.7, p = .055$ **	$\chi^2(1) = 32, p = 1.5e\text{-}08$ ***	$\chi^2(1) = 17, p = 3.8e\text{-}05$ ***
TWO	$\chi^2(1) = 47, p = 5.5e\text{-}12$ ***	$\chi^2(1) = 8.8, p = .003$ **	$\chi^2(1) = 44, p = 2.8e\text{-}11$ ***
PLURAL	$\chi^2(1) = 43, p = 4.6e\text{-}11$ ***	$\chi^2(1) = 1.4, p = .24$	$\chi^2(1) = 43, p = 5.5e\text{-}11$***

Table 11. Three readings for target items (dual task experiment)

dual task	**LIT:** 'false' vs. 'literal'	**PSI:** 'literal' vs. 'primary'	**SSI:** 'primary' vs. 'secondary'
All items	$\chi^2(1) = 48\ p = 4.6e\text{-}12$ ***	$\chi^2(1) = 21, p = 4.7e\text{-}06$ ***	$\chi^2(1) = 46, p = 1.2e\text{-}11$ ***
SOME	$\chi^2(1) = 20, p = 7.9.e\text{-}06$ ***	$\chi^2(1) = 42, p = 1.2e\text{-}10$ ***	$\chi^2(1) = 35, p = 4.2e\text{-}09$ ***
ALMOST	$\chi^2(1) = 78, p < 2.2.e\text{-}16$ ***	$\chi^2(1) = 23, p = 1.9.e\text{-}06$ ***	$\chi^2(1) = 8.8, p = 0.003$ **
TWO	$\chi^2(1) = 50, p = 2e\text{-}12$ ***	$\chi^2(1) = 3.4, p = .07$	$\chi^2(1) = 21, p = 3.7e\text{-}6$ ***
PLURAL	$\chi^2(1) = 58, p = 2.4e\text{-}14$ ***	$\chi^2(1) = 1.2, p = .26$	$\chi^2(1) = 0.84 , p = 0.36$

Figure 13. Truth value judgment task (no dual-task experiment)

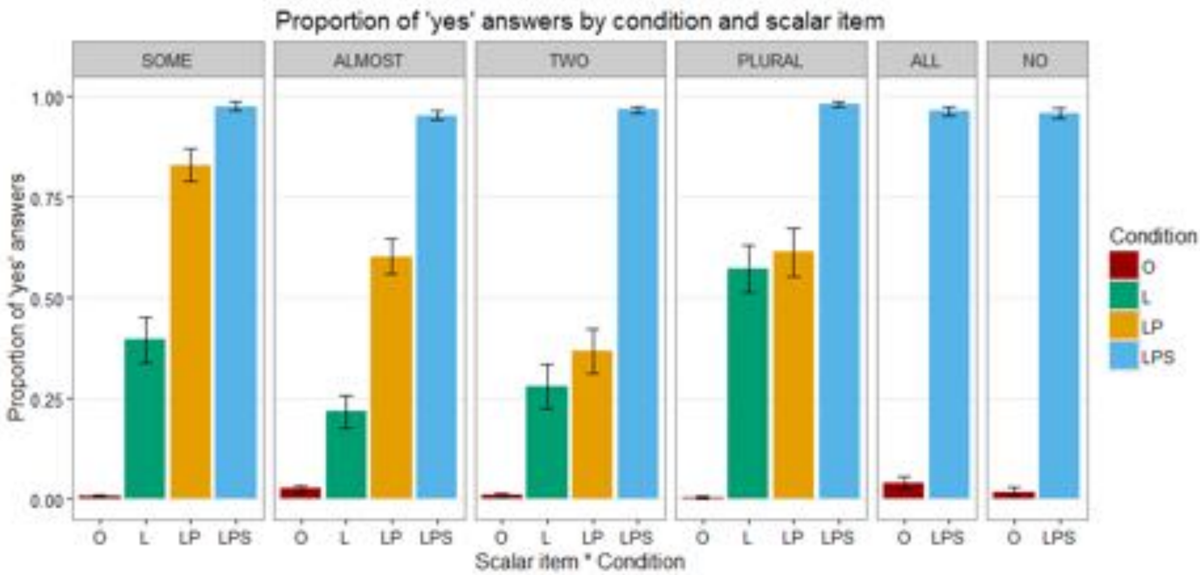
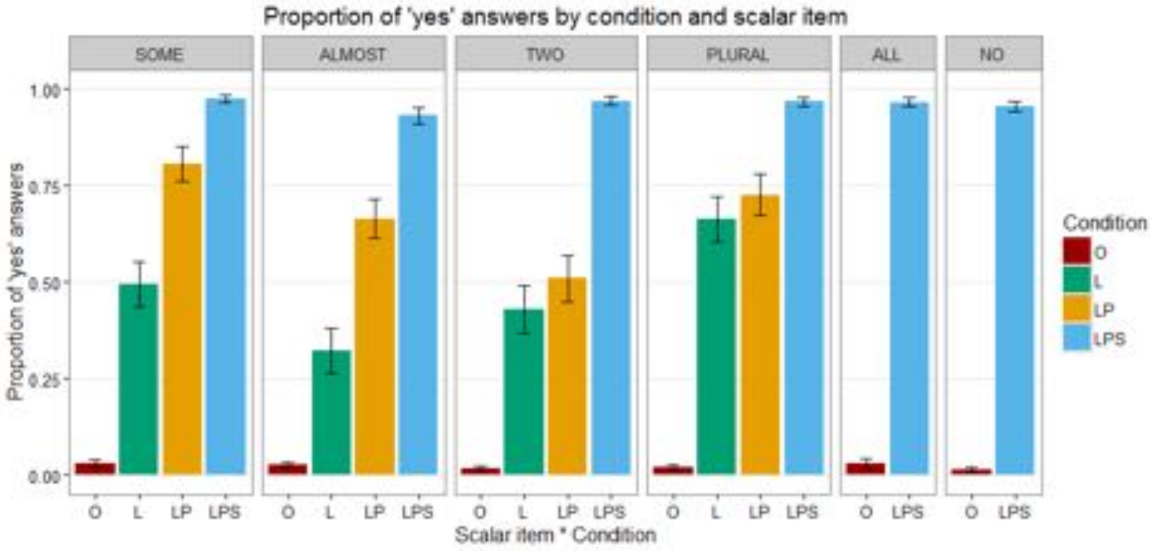


Figure 14. Truth value judgment task (dual-task experiment)



Effect of Cognitive Load

Figure 15 shows the effect of the dual-task for each scalar item, depending on the condition.

For SOME and ALMOST, the hypothesis was that the proportion of LIT would increase with Cognitive Load, following results of previous studies. For NUMERALS, the reverse pattern was expected, as found by Marty et al. (2013): the proportion of LIT would decrease when the Cognitive Load is higher. There was no precise expectation concerning PLURAL. Following the simplest neo-Gricean account (assuming that the effect is mostly due to the retrieval and manipulation of the alternative sentence), PSI and SSI should be impacted by the Cognitive Load in the same way.

First, we tested the effect of *having* a dual task on the 3 readings, comparing the no-dual-task (NO CL) to the dual-task version of the experiment (merging HIGH and LOW CL). We tested the interaction between the condition and the version of the experiment for each scalar item. Results are reported in Table 12. The effect of having a dual task is significant only for PLURAL on SSI. No general pattern emerges.

We then tested the effect of the level of Cognitive Load (HIGH vs. LOW). As summed up in Table 13, the results are not significant for PSI and SSI, whatever the condition scalar item tested.

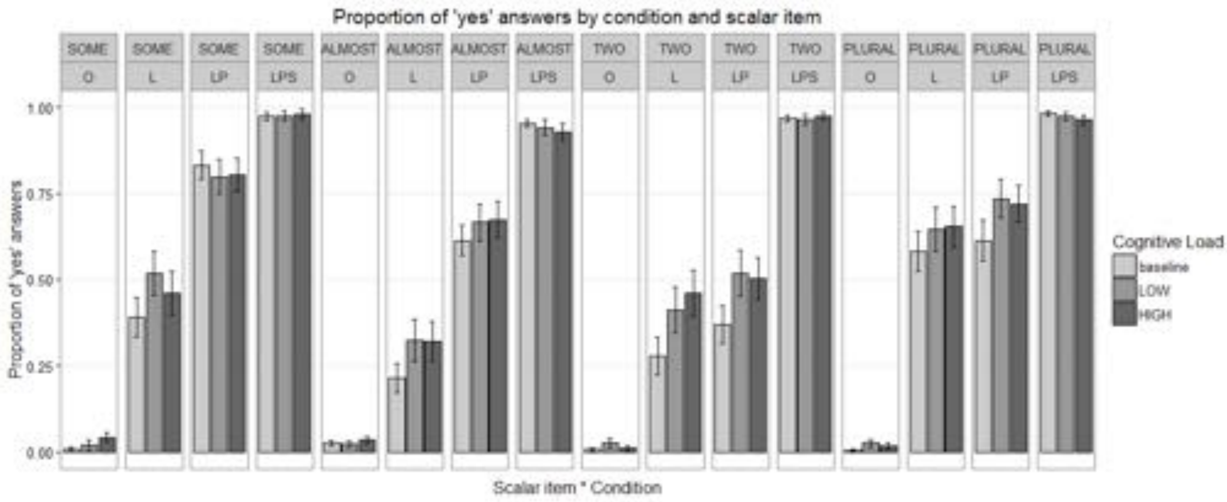
Table 12. Effect of having a dual-task for all items

dual task vs. baseline	**LIT:** 'false' vs. 'literal'	**PSI:** 'literal' vs. 'primary'	**SSI:** 'primary' vs. 'secondary'
All items	$\chi^2(1) = 0.23, p = .63$	$\chi^2(1) = 0.73, p = .39$	$\chi^2(1) = 0.53, p = .46$
SOME	$\chi^2(1) = 1.4, p = .44$	$\chi^2(1) = 0.031, p = .86$	$\chi^2(1) = 0.88, p = .35$
ALMOST	$\chi^2(1) = 1.6, p = .23$	$\chi^2(1) = 0.02, p = .89$	$\chi^2(1) = 53, p = .47$
TWO	$\chi^2(1) = 1.8, p = .18$	$\chi^2(1) = 1.6, p = .21$	$\chi^2(1) = 0.52, p = .47$
PLURAL	$\chi^2(1) = 0, p = 1$	$\chi^2(1) = 0.56, p = .46$	$\chi^2(1) = 5, p = .025*$

Table 13. Effect of the cognitive load for all items

LOW vs. HIGH load	**LIT:** 'false' vs. 'literal'	**PSI:** 'literal' vs. 'primary'	**SSI:** 'primary' vs. 'secondary'
All items	$\chi^2(1) = 3e\text{-}4, p = .99$	$\chi^2(1) = 2e\text{-}04, p = .99$	$\chi^2(1) = 0.44, p = .51$
SOME	$\chi^2(1) = 2.9, p = .09$	$\chi^2(1) = 0.17, p = .68$	$\chi^2(1) = 0.08, p = .78$
ALMOST	$\chi^2(1) = 0.23, p = .64$	$\chi^2(1) = 0.002, p = .96$	$\chi^2(1) = 0.04., p = .84$
TWO	$\chi^2(1) = 2.1, p = .15$	$\chi^2(1) = 1.9, p = .16$	$\chi^2(1) = 3.4, p = .06$
PLURAL	$\chi^2(1) = 0.44, p = .51$	$\chi^2(1) = 0.87, p = .35$	$\chi^2(1) = 0.22, p = .64$

Figure 15. Effect of Cognitive Load on the answer, depending on the scalar item



Conclusion
Given that findings of previous studies regarding the effect of Cognitive Load on the computation of SI (De Neys & Schaeken, 2007; Marty & Chemla, 2011; Marty et al., 2013) are not replicated, and statistical evidence being arguably weak especially when "multiple comparisons" are taken into account, few conclusions can be drawn from the experiment.

Comparing the dual-task version to the baseline with no dual-task, the results go in the expected direction, even if they overall turn out not to be significant. However, comparing a dual-task and a no-dual-task experiment does not directly inform on the involvement of memory resources, and this cost could be due, for example, to the effect of switching between two tasks instead. It remains to be discussed how that type of interference could affect the derivation.

There are several explanations for the fact that we do not replicate previous results: First, it is possible that our task was too easy: even if the results on the Letter Memory Task alone showed that 4-letters-sequences were more demanding than 2-letters-sequnces, the overall rate of correct answers on the Memory Task is quite high. The difference between the two levels of Cognitive Load was perhaps not strong enough. Second, contrary to Marty et al., this was an online experiment, which meant that there were factors not controlled for. Last, another factor differed between our experiment and Marty et al.'s: the linguistics task. Marty et al. used a graded judgment task, possibly more difficult than ours.

Appendix D. Instructions

Instructions for Experiment 1 and 2

Peter and Mary are playing a card game. At each round, eight cards are put on the table. Some cards can been seen by both Peter and Mary, and some other cards can be seen only by Peter (or only by Mary).
After they have looked at the cards, Peter (or Mary) makes a statement about the cards.
Your task is to indicate if Peter (or Mary) could say what he (or she) said, on the basis of his (or her) information.
Here are some examples:

1: Mary: "All of the cards are spades."
Can Mary say that?
NO, because even if in fact the sentence is true, she does not have enough information to say that.

2: Peter: "All of the cards are hearts."
Can Peter say that?
NO, because this is false (and he has enough information to know that it's false).

3: Mary: "All of the cards are clubs."
Can Mary say that?
YES, because she can be sure the sentence is true.

<u>Training with feedback</u>
Five non ambiguous sentences were used, listed above. The order of the sentences was randomized.

(1)	"All of the cards are [hearts]."	(3)
(2)	"The first card is a [heart]."	(3)
(3)	"Fewer than five cards are [hearts]."	(3)
(4)	"There is the same number of [hearts] and [spades]."	(3)
(5)	"The last card is a [heart]."	(2)

7 were attributed to Mary, 7 to Peter. 5 expected answer: yes; 5 expected answer: *no* because of not enough information; 3 expected answer: *no* because false. In the control experiment, as we used sentences with FIRST and LAST, we changed examples (4) and (5).

Instructions for the Dual Task Experiment

<u>Instructions (1/2) (Truth value judgment task)</u>
This part of the instructions was the same than in experiment 1.

<u>Training with feedback</u>: Four non ambiguous sentences were used, listed above. The order was randomized.

(1) "All of the cards are hearts."
(2) "The second card is a club."
(3) "Fewer than five cards are hearts."
(4) "There is the same number of spades and diamonds."
(5)

<u>Instructions (2/2) (Letter Memory Task)</u>
That's not all:
Before each of these questions, you will be shown random letters.
Remember them: after you have seen the cards and given your answer, you will be asked to reproduce the same sequence of letters in reverse order.
For example, you may see ABCD, then you will answer a question about a round of cards, and then you will be asked to reproduce the sequence of letters in reverse order, here: DCBA.
Please give your answer IN CAPITAL LETTERS, without leaving any space between the letters.
It is very important that you memorize correctly these letters: stay focused!

<u>Training with feedback</u>: The four same sentences were used.
The number of letters presented in the training sequence depended on the order of the blocks for the participant (LOW CL - HIGH CL vs. HIGH CL – LOW CL).

Appendix E. Results of statistical analysis comparing the different instantiations of conditions 'false' and 'primary' (Experiment 1) and 'false', 'primary' and 'secondary' (Experiment 2).

Table 14. Differences between the different instantiations of conditions 'false' and 'primary', for each scalar item (Experiment 1)

	'primary-ign-false' vs. 'primary-ign-ambig'	'false-ign-false' vs. 'false-ign-true'	'false-kn' vs. 'false-ign-false'	'false-kn' vs. 'false-ign-true'
SOME	$\chi^2(1) = 2.85, p = 0.091$	$\chi^2(1) = 1e\text{-}04, p = 0.994$	$\chi^2(1) = 6.75, p = 0.0093$**	$\chi^2(1) = 10.1, p = 0.0015$**
ALMOST	$\chi^2(1) = 0.108, p = 0.741$	$\chi^2(1) = 0.099, p = 0.751$	$\chi^2(1) = 6.58, p = 0.01027$*	$\chi^2(1) = 0.0239, p = 0.877$
TWO	$\chi^2(1) = 0.216, p = 0.640$	$\chi^2(1) = 1, p = 1$	$\chi^2(1) = 0.0026, p = 0.959$	$\chi^2(1) = 2.26, p = 0.132$
PLURAL	$\chi^2(1) = 0.324, p = 0.568$	$\chi^2(1) = 1.98, p = 0.159$	$\chi^2(1) = 6.84, p = 0.0089$**	$\chi^2(1) = 2.17, p = 0.140$
All items	$\chi^2(1) = 0.229, p = 0.632$	$\chi^2(1) = 2.18, p = 0.139$	$\chi^2(1) = 2.00, p = 0.157$	$\chi^2(1) = 2.61, p = 0.105$

Table 15. Differences between the different instantiations of conditions 'false', 'primary' and 'secondary', for each scalar item (Experiment 2)

	'secondary-kn' vs. 'secondary-ign'	'primary-ign-false' vs. 'primary-ign-ambig'	'false-ign-false' vs 'false-ign-true'	'false-kn' vs 'false-ign-false'	'false-kn' vs 'false-ign-true'
SOME	$\chi^2(1) = 0.44, p = 0.51$	$\chi^2(1) = 0.58, p = 0.45$	$\chi^2(1) = 0.047, p = 0.8$	$\chi^2(1) = 0.09, p = 0{,}76$	$\chi^2(1) = .189, p = 0.66$
ALMOST	Not Applicable	$\chi^2(1) = 0.03, p = 0.86$	$\chi^2(1) = 0.72, p = 0.39$	$\chi^2(1) = 0.903, p = 0.34$	$\chi^2(1) = 0.036, p = 0.85$
TWO	Not Applicable	$\chi^2(1) = 0.21, p = 0.65$	$\chi^2(1) = 1.38, p = 0.24$	$\chi^2(1) = 4.13, p = 0.042$*	$\chi^2(1) = 0.90, p = 0.34$
PLURAL	$\chi^2(1) = 0.45, p = 0.49$	$\chi^2(1) = 0.39, p = 0.53$	$\chi^2(1) = 0.22, p = 0.64$	*Fail to converge*	$\chi^2(1) = 0.08, p = 0.77$
All items	$\chi^2(1) = 0.22, p = 0.64$	$\chi^2(1) = 0.02, p = 0.88$	$\chi^2(1) = 0.039, p = 0.84$	$\chi^2(1) = 0.006, p = 0.93$	$\chi^2(1) = 0.008, p = 0.93$