

# Priming Reveals Similarities and Differences between Three Purported Cases of Implicature: Some, Number and Free Choice Disjunctions<sup>☆</sup>

Marie-Christine Meyer<sup>a,\*</sup>, Roman Feiman<sup>b</sup>

<sup>a</sup>*Leibniz-Center for General Linguistics (ZAS), Berlin, Germany*  
<sup>b</sup>*Department of Psychology, University of California, San Diego, USA*

5

---

## Abstract

Across a wide variety of semantically ambiguous sentences, implicature has been proposed as a single mechanism which can derive one reading from another in a systematic way. While a single formal mechanism for computing implicatures across disparate cases has an appealing parsimony, differences in behavioral and processing signatures between cases have created a debate about whether the same computation really is so widely shared. Building on previous work by Bott & Chemla (2016), three experiments use structural priming to test for shared computations across three purported cases of implicature: the quantifier *some*, number words, and Free Choice disjunctions. While we find evidence of a shared computation between the enriched readings of *some* and number words, we find no evidence that Free Choice readings involve any shared computation with either *some* or number. Along with evidence of a shared mechanism between *some* and number implicatures, we also find substantial differences between these two cases. We propose a way to reconcile these findings, as well as seemingly contradictory prior evidence, by understanding implicature as a sequence of separable sub-computations. This implies a spectrum of possibilities for which sub-computations might be shared or distinct between cases, instead of a single implicature mechanism that can only be either present or absent.

*Keywords:* Psycholinguistics, Pragmatics, Semantics, Implicature, Priming, Free Choice, Bare Numerals

---

```
;;echo=FALSE, cache=TRUE;_j_ = Custom functions for printing Chisq and p-values in  
apa format Define methods for p-values custom.print.p.value.Chi _j_ function(x,y, ...)  
ifelse(x[y,"Pr(_j_Chisq)"]_j_=0.001, paste(sprintf("=
```

---

<sup>☆</sup>For comments and discussion, we would like to thank audiences at SALT30, UCSD, Leibniz-ZAS Berlin, the Crete Summer School of Linguistics, the Syntax-Semantics Seminar at the University of Nantes, and three anonymous reviewers. We would also like to thank Marlena Jakobs for help with the experimental stimuli. Part of this work has been supported by DFG grant SA 925/11-2 within the *XPrag.de* priority program (SPP 1727).

\*Corresponding author

*Email addresses:* [macrst@alum.mit.edu](mailto:macrst@alum.mit.edu) (Marie-Christine Meyer), [roman\\_feiman@brown.edu](mailto:roman_feiman@brown.edu) (Roman Feiman)

*Preprint submitted to Journal of Memory and Language*

```

custom.print.p.value.sum.glmer j- function(x,y, ...) ifelse(x[y,"Pr(i—z—)"]<=0.001,
paste(sprintf("=
Define method for t- and z-statistics custom.print.2.dec.Chi j- function(x,y, ...) paste(sprintf("custom.print.2.de
j- function(x,y, ...) paste(sprintf("
Define a method to take a summary(glmer) output, then print the Z-score value and
the p-value in APA format custom.print.sum.glmer j- function(x,y, ...) sprintf("Z =
Define a method for printing DF in Anova output. custom.print.df j- function(x,y,
...) sprintf("
Define a method to take an Anova(glmer) output, then print the ChiSq value and the
p-value in APA format custom.print.Anova j- function(x,y, ...) sprintf("χ²(
Load data from Experiments Exp1 j- new.env(parent = emptyenv()) Exp3 j- new.env(parent
= emptyenv()) Comp j- new.env(parent = emptyenv())
load("Exp 1 Results for Overleaf.RData", envir = Exp1) load("Exp 3 Results for
Overleaf.RData", envir = Exp3) load("Exp 5 and 7 Combined Results for Overleaf.RData",
envir = Comp) @

```

## 1. Introduction

Since Grice first suggested that even unspoken sentences can contribute to the meaning of an utterance, half a century of research on implicatures has formalized and substantiated the idea in different frameworks. A significant body of work in this area has focused on showing how implicature can be extended to a diverse set of phenomena, far exceeding the scope of Grice’s original proposal. But just because a formal mechanism can be designed by the theorist to derive certain phenomena, does not mean that human minds actually use this mechanism in the course of language comprehension, especially when alternative explanations for the same phenomena are readily available.

In this paper, we apply psycholinguistic tools to examine which – if any – components of a hypothesized general implicature mechanism might be shared across different categories of utterances. We start by noting that implicature consists of several sub-computations. Focusing on these sub-computations will allow us to develop a finer-grained view of implicature: Rather than asking whether or not implicature *as a whole* is applied across different categories, we can ask which of the sub-computations are shared across categories. Indeed, as our experimental results reveal, the sub-computations are not all equally likely to be applied in different categories, necessitating a finer-grained theoretical notion of implicature, which we will develop in this paper.

Using the structural priming method, we study three different domains: the quantifier *some*, number words, and Free Choice disjunctions. In each case, two readings are available; for *some* and number these are shown below:

- (1) Some of the houses have a fence
  - a. Basic reading: At least some of the houses have a fence
  - b. Enriched reading: Some but not all of the houses have a fence
- (2) There are four dogs
  - a. Basic reading: There are at least four dogs
  - b. Enriched reading: There are exactly four dogs

We refer to the reading that is hypothesized to be the input to the implicature mechanism as the *basic* reading, and to the reading that is hypothesized to be the output of the implicature mechanism as the *enriched* reading, following Bott & Chemla (2016), among many others. While this terminology transparently reflects the hypothesis we set out to test, it does not reflect any prior commitment to the truth of that hypothesis.

### 1.1. The implicature mechanism for deriving enriched readings

Theories of implicature posit a general mechanism that takes the basic, lexically encoded reading of a sentence as input and yields its enriched reading as output. We home in on the following two sub-computations, assumed by all accounts of implicature:

- (3) Core Implicature Mechanism
  1. Generate alternative(s)
  2. Negate alternative(s)

The alternatives are sentences which were not asserted, but would have been equally relevant and more informative in the given context. The generation of alternatives is that computation which yields those alternatives; it can itself be broken down into several components (Katzir, 2007). For *Some of the houses have a fence*, it involves retrieving specific lexical material (the quantifier *all*) and replacing the lexical item *some* in the original sentence, generating the alternative sentence *All of the houses have a fence*. The alternatives are then negated; for our example, conjoining the negation with the original sentence yields *Some of the houses have a fence & not all of the houses have a fence* as the final output – the enriched reading of the sentence.

The enriched reading of sentences like *There are four dogs* has been argued to be derived in much the same way (e.g. van Rooij & Schulz, 2006; Horn, 1989; Gazdar, 1979). According to this implicature-based view of numerals, the basic, lexically stored reading of *four* can be paraphrased as *at least four* or *four or more*. On this reading, *There are four dogs* would be true in a scenario in which there are in fact six dogs. But a more natural reading is of course the *exactly*-reading, on which the sentence is judged false in this scenario. By hypothesis, this reading is derived through the same implicature mechanism as in the case of *some*. First, generating the alternative will again involve retrieving a specific lexical alternative, in this case the number word *five*, followed by appropriate replacement in the original sentence, yielding the alternative sentence *There are five dogs*. Next, this alternative is negated. When conjoined with the original sentence, the enriched *exactly*-reading is obtained: *There are (at least) four dogs & it's not the case that there are (at least) five dogs*. In this way, the hypothesized mechanism underlying the enriched readings of number and *some* sentences differs only in the specific alternatives that are generated and then negated.

The hypothesis that the enriched readings of different sentences are derived by a general implicature mechanism has the strong appeal of parsimony, extending far beyond the domain of quantifiers and number words. Implicature has been invoked to explain other cases of seeming ambiguities in logical connectives, as in *John bought apples or oranges* (enriched: *John bought apples or oranges, but not both*), gradable adjectives, as in *The soup is warm* (enriched: *The soup is warm, but not hot*), modal expressions, as in *John may come* (enriched: *John may come, but doesn't have to*), and even attitude

verbs, as in *John believes/thinks that it's ok* (enriched: *John thinks, but doesn't know, that it's ok*; see van Tiel et al. 2014; Hirschberg 1991 for an overview).

Importantly, the question of whether a shared implicature mechanism is responsible for these disparate phenomena is independent of many active debates concerning the nature of implicature. (Neo-)Gricean (Geurts, 2010; van Rooij & Schulz, 2004; Hirschberg, 1991; Grice, 1989, 1975; Gazdar, 1979; Horn, 1972), grammatical (Meyer, 2013; Fox, 2007; van Rooij & Schulz, 2006; Chierchia, 2004), relevance-theoretic (van Roooy, 2002; Merin, 1999; Carston, 1998; van Kuppevelt, 1996; Sperber & Wilson, 1986), game-theoretic (Franke, 2011; Benz, 2006; Parikh, 2001), and Bayesian (Goodman & Frank, 2016; Bergen et al., 2016; Potts et al., 2015) theories disagree on whether or not implicatures are computed as part of the grammar or as a pragmatic inference. They also differ on the question of what causes implicatures to be computed in the first place: Gricean maxims of cooperative conversation, domain-general principles of rational behavior, or something else still. What matters for our present purpose is that, with the exception of probabilistic theories,<sup>1</sup> all of these theories share the core assumption that when implicatures are computed, the mechanism involves the two sub-computations outlined in (3) above (see also Bott & Chemla 2016, for discussion). It is this shared assumption which we will be investigating in this paper.

### 1.2. Free Choice readings as (recursive) implicatures

Along with the quantifier *some* and number words, we investigate the case of Free Choice disjunctions (for review see Meyer 2020). Alonso-Ovalle (2006) and Fox (2007) first showed how to extend the implicature mechanism to account for the ambiguity of disjunctions like the following:

- (4) John can take the red umbrella or the yellow umbrella

Under one reading, which we will call the *uncertainty* reading, the sentence is used as a report on John's options from a speaker who has incomplete knowledge of what these options are. The uncertainty reading is derived by the standard assumption that the lexical meaning of *or* can be modeled by the logical connective  $\vee$ . This predicts (4) to be true in three scenarios: (i) if John can take the red umbrella, but not the yellow one, (ii) if John can take the yellow umbrella, but not the red one, and (iii) if John can take both umbrellas.<sup>2</sup>

The Free Choice (FC) reading, on the other hand, conveys that John has *both* the option of taking the red umbrella and the option of taking the yellow one, though he

---

<sup>1</sup>Probabilistic theories model the step from the basic to the enriched reading as a probabilistic, and more specifically, a Bayesian inference about the likelihood of a reading given the speaker's utterance (e.g., what is the likelihood that the speaker who said *Some of the houses have a fence* meant to convey that all houses have a fence? see Goodman & Frank 2016; Bergen et al. 2016; Potts et al. 2015). While these accounts do rely on the presence of a linguistic alternative to the uttered sentence, enrichment takes the form of a probabilistic inference about the truth or falsity of the alternative. Therefore, regarding the core mechanism given in (3) above, the negation of alternatives is not necessarily part of the relevant mechanism.

<sup>2</sup>As is well-known, scenario (iii) is often ruled out independently. In an implicature-based account, this could be derived by negating the alternative *John can take the red umbrella and the yellow umbrella*, which is generated by substituting *and* for *or* in the original sentence. This meaning component is independent of the distinction between the uncertainty and the FC reading, and can be derived independently for both meanings, as discussed in e.g. Fox (2007).

may not necessarily have the option of taking two umbrellas at once. Under this FC reading, the speaker has full knowledge of all options, and might even be the one granting permission to choose between them (as in a parent saying, *You may have cake or ice-cream for dessert*).

How could an implicature mechanism derive the FC reading of (4) from its hypothesized basic uncertainty reading? In the cases of *some* and number, the hypothesized implicature mechanism involves the negation of the *basic* reading of alternatives. To derive FC readings, it has instead been proposed that the alternatives are first themselves enriched, and it is the *enriched* readings of the alternatives which are then negated (Fox, 2007; Alonso-Ovalle, 2006; Kratzer & Shimoyama, 2002). This is schematically illustrated below (the parts that differ from (3) are underlined):

- (5) Recursive implicature mechanism for a sentence *can [A or B]*
1. Generate alternatives
  2. Compute the enriched meanings of the alternatives
  3. Negate the enriched meanings of the alternatives

For FC sentences like (4), the proposal is that the alternatives in question are the disjuncts, *John can take the red umbrella* and *John can take the yellow umbrella*. The enriched meanings of these two alternatives are derived by conjoining each of them with the negation of the other (which means they are alternatives to each other). This yields their enriched meaning: *John can take the red umbrella & not the yellow umbrella* and *John can take the yellow umbrella & not the red umbrella*. This concludes the sub-computation in step 2 above. In a last step, these enriched alternatives are negated, and conjoining them with the basic uncertainty reading yields the FC reading of the whole disjunction:

- (6) John can take the red umbrella or the yellow umbrella
- a.  $\& \neg$  (John can take the red umbrella and not to the yellow one)
  - b.  $\& \neg$  (John can take the yellow umbrella and not the red one)
- = John can take the red umbrella and he can take the yellow umbrella

Because the alternatives hypothesized to be negated in the derivation of the FC meaning are themselves enriched, the mechanism can be characterized as *recursive*: it takes the output of an implicature computation as its input.

The mechanism may be ornate, but the resulting parsimony is appealing, as the enriched readings of sentences with quite disparate content – number words, *some*, and disjunctions combined with modals – can all be derived by the application of one core mechanism, a procedure for generating and negating alternatives. Positing a category-general implicature mechanism also makes a clear, testable prediction. If, whenever the enriched meaning is accessed, an implicature computation consisting of the sub-computations in (3) above is performed, then any processing signature of these computations should be shared across otherwise very distinct categories of enriched meanings.

This prediction has been tested extensively, but the results have not been conclusive. Many findings show that the different cases for which implicature-based accounts have been proposed exhibit substantial differences. Before we go on to the present study, we briefly review the most relevant prior results.

### 1.3. Previous Experimental Evidence

The evidence that implicature is used to derive the enriched meaning of *some* has been mixed. On the one hand, several studies of online sentence processing with adults have found that the basic reading of *some* is accessed earlier, and verified faster, than the enriched reading in comprehension. This is consistent with the enriched reading being the output of an additional implicature computation that is not involved in accessing the basic reading (e.g. Huang & Snedeker, 2018; Tomlinson et al., 2013; Huang & Snedeker, 2011, 2009; De Neys & Schaeken, 2007; Bott & Noveck, 2004).

But other experimental studies have produced contradictory results. For example, Degen & Tanenhaus (2015) replicate the finding of higher reaction times for the enriched vs. basic reading of *some*, but show that these results might be due to decreased acceptability of *some* in the relevant experimental conditions. In a similar vein, Grodner et al. (2010) and Feeney et al. (2004) argue that increased processing times for enriched readings are due to independent factors rather than implicature computations. Foppolo & Marelli (2017) show that when the confound identified by Degen & Tanenhaus (2015) is controlled for, processing times of the enriched vs. basic reading do not differ (see also Breheny et al. 2013). These findings do not support the hypothesis that the enriched reading is derived from the basic one via a cognitively costly implicature computation, but are more consistent with accounts in which the enriched reading is stored directly as part of a lexical ambiguity of the quantifier *some*. Still, the debate continues, with Huang & Snedeker (2018) challenging these conclusions and reporting further evidence consistent with multi-step processing of the enriched meaning of *some*.

The empirical evidence on whether implicature generates *exactly*-readings of number words is likewise mixed, with more evidence against this hypothesis than in the case of *some*. First, not only can the enriched readings of number words be accessed as quickly as their basic readings, Huang & Snedeker (2011, 2009) also find that in the very same contexts in which hearers are slow to access the enriched reading of *some*, they access the enriched reading of *two* as quickly as both the basic reading of *some* and the meaning of *all*, which is not derived via implicature on any account (Huang & Snedeker, 2011). Huang et al. (2013) also show that hearers access the basic reading of *some* more often than the basic reading of number when given the choice (see also Marty et al. 2013). Lastly, Marty et al. (2013) and De Neys & Schaeken (2007) found that, while increased cognitive load made the basic reading of *some* more accessible, it instead made it *harder* to access the basic, *at least* reading of number words. These results are consistent with the possibility that implicature underlies the enriched reading of *some*, but not of number words.

Weighing on the other side, there is also evidence of a multi-step implicature mechanism generating the enriched reading of number words. Results from Panizza et al. (2009) suggest that the choice between the enriched and the basic reading correlates with logical properties of the linguistic environment for both *some* and numbers. The enriched meaning is preferred for both items in upward-entailing environments, where it would make the whole sentence more informative overall, and dispreferred in downward-entailing environments, where it would make the whole sentence less informative. Taken alone, the significance of this type of evidence is limited by its correlational nature, and the relative preference of basic vs. enriched readings of two different items in the same environment can be explained by a general preference for an overall more informative

statement, regardless of whether or not the enriched readings are generated by the same mechanism in both items. The stronger evidence from Panizza et al. is that in upward-entailing contexts in which enriched readings are more informative, reading times for number words were higher than in downward-entailing contexts, in which basic readings would be more informative and therefore expected to be preferred. This shows that the two readings are not equivalent when controlling for informativity, and more strongly suggests that the *exactly*-reading of numbers involves an additional computation (see also Guasti et al., 2005; Noveck et al., 2002; Chierchia et al., 2001; Gualmini et al., 2001).

The strongest evidence in favor of implicatures generating the enriched *exactly*-readings of number words comes from structural priming. Bott & Chemla (2016) show that when participants access the *exactly*-reading of number words, they are subsequently more likely to access the enriched reading of *some* than its basic reading, suggesting a shared mechanism responsible for generating both enriched meanings. This test has not previously been extended to FC disjunctions, one of the primary targets of our investigation below.

For FC disjunctions, experimental studies in general are sparse so far, and the most relevant study is inconclusive. If it is indeed a recursive implicature computation which derives the FC reading, accessing the FC reading should take longer compared to the uncertainty reading, analogously to the logic of studies comparing basic vs. enriched readings of both *some* and number words.<sup>3</sup> In the only experimental test of this hypothesis so far, Chemla & Bott (2014) instead found *faster* response times for FC compared to uncertainty readings. On this basis they argue that the FC reading cannot stem from an implicature computation requiring additional processing steps. There is, however, a plausible alternative explanation for this finding. Recall that the uncertainty reading of a FC sentence like *John can have the red bottle or the yellow bottle* states that he can have (at least) one of the two, but the speaker is uncertain as to which one. In a scenario in which the speaker already knows that John can have the red bottle, the whole disjunction would be infelicitous, presumably because the assumption of speaker uncertainty is violated. The items used by Chemla & Bott (2014) do not satisfy this uncertainty condition, and the basic reading is therefore infelicitous in their scenarios, though it is true in a strictly logical or semantic sense. The increased processing times compared to the FC reading could be due to participants trying to accommodate the infelicitous reading.

In sum, there is mixed evidence that both the enriched reading of *some* and number words are derived by an implicature mechanism, as proposed in the theoretical literature. The two cases also differ, with the enriched meaning of number words being relatively easier to access. Whether implicature is involved in generating FC readings is currently unclear.

---

<sup>3</sup> As stated in footnote 2 above, the FC reading can have an additional, optional scalar implicature, namely, that both options are not available at the same time (e.g., John cannot have both the yellow and the red bottle). This independent implicature is thus expected to add to processing times of both readings, if present, and may make it more difficult to find a difference between the two (if, for instance, additional implicatures are less costly). This is one reason why we do not look for a processing time signature in this case. To further control for this independent factor, both our study and Chemla & Bott (2014) only use scenarios in which this additional component is kept constant, by making it irrelevant in all experimental items used to test either uncertainty or FC readings. The way this can be achieved is to provide background information to participants that there is never a possibility for both options.

#### 1.4. Overview of Experiments

To test whether there is a shared implicature computation, we will look for a causal link: does deriving an enriched reading for one sentence (e.g., a FC sentence) increase the likelihood of deriving an enriched reading for a subsequent sentence from a different category (e.g., a number sentence)? As Bott & Chemla (2016) pointed out, the majority of studies on whether different types of enriched readings involve a shared implicature mechanism are correlational, comparing independent patterns of distributional or processing data from each category. Structural priming is a method that can provide stronger, causal evidence in this ongoing debate. The underlying rationale is that if a computation or representation is involved in the processing of one sentence, it will be easier to reuse in the processing of another sentence in which it is also involved (see Branigan & Pickering 2017 for an overview).

For the cases we are interested in, this translates into the following prediction. If a general implicature mechanism underlies the enriched readings of *some*, number, and FC sentences, then accessing this reading in one sentence should increase the likelihood of accessing the enriched reading again in a subsequent sentence. Importantly, this prediction holds regardless of whether the two sentences are from the same (within-category effect) or a different category (between-category effect). This is because the assumed mechanism, which involves generating and negating alternatives, is category-general.

The only previous study that uses structural priming to investigate implicature is Bott & Chemla (2016), which serves as our starting point. Our experiments have a similar design to theirs. On both prime and target trials, participants see two pictures presented below a sentence, and are asked to choose the picture that best matches the sentence. The “priming effect” is measured by how participants’ choice on a given target trial changes depending on the kind of prime trial that preceded it. The prime trials were meant to elicit either a basic or an enriched reading, for sentences from each of the three categories of interest. The target trials following each prime trial then assessed whether participants opt for an enriched or a basic reading when given the choice. A given combination of prime and target items can be either *within-category*, where both prime and target are from the same category (e.g., number), or *between-category*, where the categories are different (e.g., FC for prime, *some* for target). Since the enriched reading is typically preferred, target items did not simply present a choice between the two, but rather a choice to accept or reject the basic reading. In keeping with the picture-matching design, this was implemented as a choice between a picture that matched the basic reading, and a picture only containing the words *Better Picture*, which allows participants to indicate that they would prefer an unspecified different picture to the depicted basic reading, following Bott & Chemla (2016) and Huang et al. (2013).

In Experiment 1, we investigate both within-category priming effects and between-category effects to establish whether, and in which cases, implicature computations might be shared. A control experiment, reported in Supplementary Materials, rules out the possibility that within-category priming in this design is due to greater similarity between pictures depicting the same readings than different readings.

Experiment 2 controls for the possibility that the prime trials that were meant to elicit basic readings actually primed participants to accept dispreferred readings more generally, independently of an implicature mechanism.

Experiment 3 replicates Experiment 1 using different picture stimuli that provide another control. These pictures were designed to be category-neutral, so that a single picture could exemplify the basic reading of a *some*, number, and FC sentence simultaneously (and likewise for the enriched readings). This ensures that differences in the priming effects between within-category and between-category trials are due to the processing of the sentences involved, rather than to e.g., basic *some* pictures being more similar to basic number pictures than to basic FC pictures.

## 2. Experiment 1

This experiment had two aims. The first was to replicate Bott and Chemla’s results for both within- and between-category priming with *some* and number, using experimental items that substantially differed from Bott and Chemla’s (see Figure 1). The second goal was to test the hypothesis that FC readings arise through recursive implicatures, by testing for priming effects between *some* and number sentences on the one hand, and FC sentences on the other. Experiment 1 thus included all three categories – *some*, number, and FC sentences – in all nine possible prime-target configurations, testing both within- and between-category priming. Each target category (number, *some*, FC) corresponded to one between-subject condition, with all three categories serving as primes manipulated within-subject.

### 2.1. Methods

*Participants.* We recruited 400 unique participants per condition (total N=1200), using Amazon’s *Mechanical Turk* platform. This sample size was chosen following the recommendations of Mahowald et al. (2016), who estimated that it corresponds to power above 0.9 for the size of main effect typically found in syntactic structural priming paradigms with a comparable number of items. Out of the 1200 participants, we excluded 19 who reported that their first language was not English. In order to ensure that participants were paying attention throughout the experiment, we also excluded participants who responded incorrectly on more than 10% of filler trials. In Experiment 1, this excluded an additional 147 participants.

*Materials.* Each item consisted of a sentence presented with two pictures. Within each sentence category, the same sentence frames were used in both prime and target items:

- (7) Sentences for *some* items  
Some of the houses have a X  
X = bike, dog, flower, tree, pond, fence, chimney, flag
- (8) Sentences for number items  
There are four Xs  
X = bike, dog, flower, tree, pond, bridge, house, car
- (9) Sentences for FC items  
He can go to the red truck or to the yellow truck

There were three kinds of pictures for each category: (i) false pictures, which did not match any reading of the sentence, (ii) basic pictures, which matched the basic readings of number and *some* sentences and the uncertainty reading of FC sentences, but were false under their respective enriched/FC readings, and (iii) enriched pictures, which matched the respective enriched/FC readings of the three categories. For example, all number sentences stated there were *four* Xs. A false picture showed *two* Xs, a basic picture showed *six*, and an enriched picture showed *four* Xs. For the *some* sentences, the false picture showed *none* of the houses having an X, the basic picture showed *all* of the houses having an X, and the enriched picture showed *some but not all* of the houses having an X.

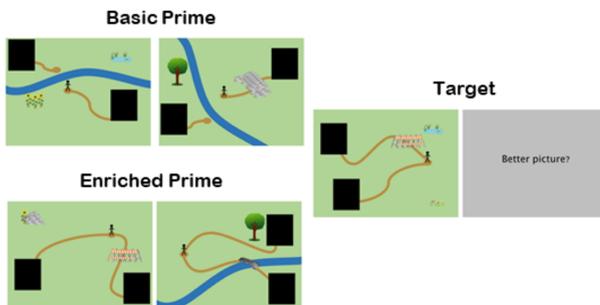
The pictures paired with FC sentences require more explanation. On the uncertainty reading (i.e., the basic reading), the sentence *He can go to the red truck or to the yellow truck* is typically accompanied by an inference that the speaker is not certain which truck he can go to (an ignorance implicature; e.g. Sauerland 2004).

To make the basic picture acceptable under this reading of the sentence, we introduced speaker uncertainty by covering all of the trucks, but not the roads which lead to them, and instructed participants that every picture had one red truck and one yellow truck, but that their locations could vary. False pictures showed a figure standing between two trucks, both covered by black boxes, with the roads to both blocked. Although participants could not tell which truck was which, the key to evaluating the sentence is that the figure cannot reach either truck. Basic pictures showed the figure with an accessible path to just one of the (covered) trucks, so participants would be uncertain about whether this was the yellow or the red truck. Enriched pictures showed the figure with accessible paths to both covered trucks, so that the figure could access both, no matter which truck was which.

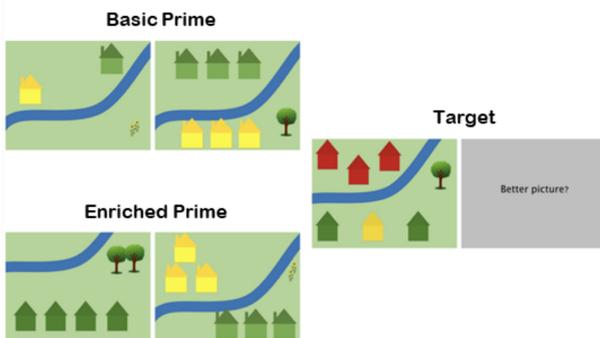
Finally, we introduced a narrator at the beginning of the experiment, whom all displayed sentences were attributed to. Participants were told that the narrator would describe all of the pictures in the experiment the way both she and the participant could see them, so that the narrator and the participant always had the exact same basis for evaluating the sentences. For the FC items, this meant that both were uncertain about which truck was which (see Appendix 1 for the script).

In all sentence categories, basic prime items presented a choice between a false picture and a basic picture. Enriched primes presented a choice between a basic picture and an enriched picture. The side each picture type appeared on was counterbalanced across prime trials within-participant. Finally, target items presented a choice between a basic picture and the *Better Picture* alternative. Following Bott and Chemla (2016), the *Better Picture* choice always appeared on the right. Figure 1 shows examples of trials in all three sentence categories.

(A) He can go to the red truck or to the blue truck



(B) There are four houses



(C) Some of the houses have a fence

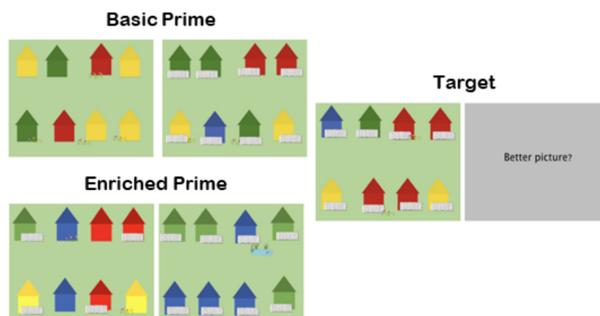


Figure 1: Sample prime and target items in Experiment 1. (A) FC items (B) number items (C) *some* items. Participants saw either a basic prime or an enriched prime item, followed by a target item. An example sentence is shown in each panel, which could be paired with either type of prime or the target item within a given sentence category. In basic prime items, the false picture is shown on the left and the basic picture is shown on the right. In enriched prime items, the enriched picture is shown on the left and the basic picture is shown on the right. The actual side of presentation of each picture on prime items was counterbalanced. On target items, the left-hand picture corresponds to the basic reading and the right-hand picture shows the 'Better Picture' option. The sides of presentation was fixed for all target items.

A full experimental *trial* is a sequence of two primes and one target. There were 8 experimental trials (4 with enriched primes, and 4 with basic primes) per target category (*some*, number, FC), for a total of 24. To minimize the effect of specific item combinations, we created two lists, such that all of the experimental trials in one list used a different combination of specific prime and target items than in the other.

There were also 72 filler items (24 from each category), randomly interspersed with experimental trials, which were meant both to make the experimental manipulation less obvious and to serve as control items, ensuring participants were engaged, read the sentences, and responded correctly to trials which had a clearly correct answer. There were three types of fillers (8 items of each type) for each of the *some* and number categories, and 2 for the FC category.

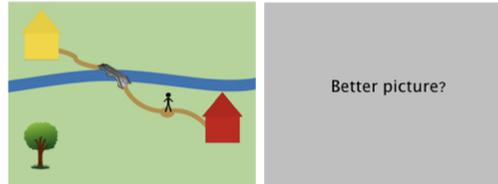
For number fillers, all sentences were “There are six Xs”. Type I fillers pitted a false picture against a *Better Picture* option. Type II fillers pitted a basic picture against a *Better Picture* option (similar to number target items), and Type III fillers pitted a basic picture against an enriched picture (similar to enriched primes).<sup>4</sup> For *some* fillers, all sentences were “All of the houses have an X”. Type I fillers pitted a false picture against a *Better Picture* option. Type II fillers pitted a true picture against a *Better Picture* option. Type III fillers pitted a false against a true picture (some-but-not-all houses versus all houses having an X). FC fillers used the sentence “He can go to the red truck, but not the yellow truck” (with the colors’ order of mention counterbalanced across trials). The trucks were uncovered, to help participants keep in mind what they had learned during the practice phase, i.e., that there was always a red and a yellow truck, even if on other trials these were not visible. Type I Fillers presented a choice between a false and a *Better Picture* option, and Type II fillers presented a choice between a true and a false picture (with correct side counterbalanced, but *Better Picture* always on the right, as in target trials). For examples, see Figures 2. See Appendix 1 for a full list of fillers.

---

<sup>4</sup>While the similarity between Type III number fillers and enriched prime items means they could act as a prime, the fillers never appeared in between actual prime and target items, since as stated earlier these were grouped into fixed prime-prime-target sequences constituting one experimental trial. It is nevertheless possible that they may have added a small additional effect, strengthening the priming of enriched number trials relative to others. This was fixed for Experiment 3.

### FC Filler Items I

He can go to the red house but not to the yellow house



### FC Filler Items II

He can go to the yellow house but not to the red house

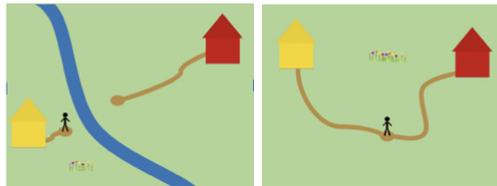


Figure 2: Examples of FC Filler Items. In Type I fillers, the choice is between a false picture (left) and a 'Better Picture' option (right). In Type II Filler Items, the choice is between a true picture (left) and a false picture (right).

*Procedure.* All of the experiments were presented using the Testable platform (<https://www.testable.org>). Participants first filled out a demographic questionnaire, which included a question about their native language. Before the start of experimental trials, participants completed a practice phase, which introduced them to the narrator as the viewpoint behind all of the written sentences they would see. The narrator "told" participants she would truthfully describe one of the two pictures presented, and the participant's job would be to guess which picture matches the description. The narrator emphasized that she can only see the same view as the participant, so while she does not always have full information, she will describe what she knows.

Participants were then given a series of practice items where they matched a sentence to one of two pictures. Each item displayed a sentence centered above two pictures, or one picture and the *Better Picture* box. Participants were instructed that, when a *Better Picture* option was present, they should choose it if they felt that the other picture did not capture the meaning of the sentence well enough. Participants made their choice by clicking on a picture, or on the *Better Picture* option. They could not click on both pictures and could not proceed to the next trial until they made a selection. Where two pictures were presented, the choice in these practice items was always unambiguous

between a false and a true picture. Several items gave participants practice responding to “covered truck” pictures, similar to what they would see on prime and target FC items in the experimental phase (see Appendix 1). Following the practice phase, participants were presented with the experimental and filler trials in randomized order. An example presentation of the *some*-target condition can be found at <https://www.testable.org/experiment/662/202179/start>.

## 2.2. Results

### 2.2.1. Analysis strategy

In addition to the participant exclusions listed above, we also excluded individual trials on which participants spent less than 500 msec, reasoning that it would not be possible to read the sentences and consider the pictures so quickly. This excluded an additional 210, which was less than 1% of all trials. Following Bott and Chemla (2016), Raffray and Pickering (2010), and others, we also analyzed only those target responses that followed a ‘correct’ response on both of the immediately preceding primes. If a participant answered ‘incorrectly’ on a prime item (i.e., chose the false picture in the basic prime, or the basic picture in the enriched prime), it could be because they did not access the reading that this prime was meant to elicit, in which case that reading could not possibly be primed.

We analyzed the data in R (v3.6.4; R Core Team 2017) using the lme4 package (v1.1.21; Bates et al. 2015) to build logit mixed-effects models. We followed the model selection strategy recommended by Matuschek et al. (2017), starting from a maximal model and using log-likelihood ratio tests to compare this to models with reduced random parameter structures, stopping upon finding a significant decrease in model fit. Rather than choosing the most reduced of these models in each case, which would have resulted in inconsistent and incomparable models across otherwise similar analyses, we selected the ‘lowest common denominator’ reduced model across similar analyses. In each case we used log-likelihood ratio tests to verify that there was no significant difference in model fit between these and the most maximal converging model. For estimates of the significance of main effects, we derived p-values from Type II Wald  $\chi^2$  tests comparing minimally different models with and without that effect. Estimates of simple effects were obtained from significance tests on coefficients of dummy coded contrast variables. All data, reproducible analysis code, and exact model specifications are available at <https://osf.io/ukrgv/files/>.

### 2.3. Accessibility of Basic and Enriched readings across sentence categories

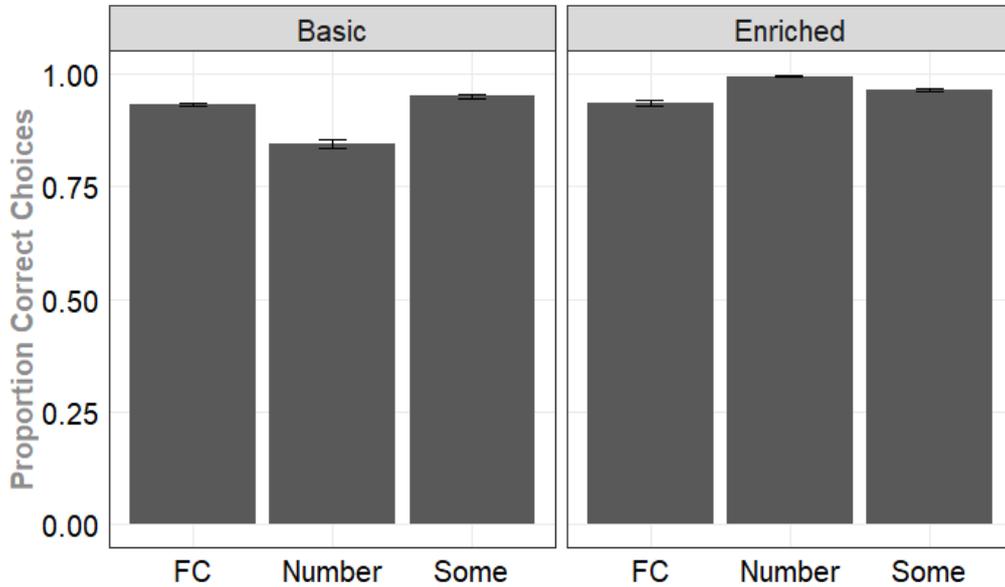


Figure 3: Responses on prime items in Experiment 1. Mean proportion of picture choices corresponding to the type of prime (basic or enriched). Error bars indicate +/- 1 standard error.

Figure 3 shows the results from basic and enriched primes for each of the three categories, reflecting the accessibility of the basic and enriched readings. ‘Correct’ responding here refers to the choice of the basic (rather than the false) picture on basic primes, and the choice of the enriched (rather than the basic) picture on enriched primes.

To test whether the accessibility of each reading differed between categories, we looked at the effect of category within basic and enriched primes in two separate analyses. In each case we set the category with the intermediate level of enriched responses as the dummy-coded reference level in order to compare differences from other categories in both directions. In enriched primes, we found that participants preferred the enriched reading of *some* sentences more than that of FC sentences ( $Z = 15.92$ ,  $p < 0.001$ ), and the enriched reading of number sentences more than that of *some* sentences ( $Z = -6.59$ ,  $p < 0.001$ ).

While enriched primes tested whether participants preferred one reading to another, basic primes tested whether participants would consistently reject a picture that was false on either reading. While this might lead to more ‘correct’ responses on basic primes, it is also possible that for some sentence categories, the basic reading is so strongly dis-preferred that it becomes almost unavailable. In that case, participants may have trouble choosing between the false and the basic picture on basic prime items. Indeed, constructing the same model only for basic prime trials, but taking the intermediate FC category as the reference level, we found that basic pictures were less consistently chosen for number sentences than for FC sentences, ( $Z = 3.90$ ,  $p < 0.001$ ), and for FC sentences

less than for *some* sentences ( $Z = -14.20$ ,  $p < 0.001$ ). The *exactly*- interpretation of number was so strongly preferred that participants rejected the *at least*- reading 15% of the time.

#### 2.4. Priming effects

Figure 4 shows the results of the priming manipulation, broken down by prime and target categories. To investigate priming effects both within- and between-category, we first fit a model with fixed effects of target category (*some*, number, or FC), category match (within or between) and prime type (basic or enriched), and all interactions. We found that each interaction was highly significant, including the three-way interaction ( $\chi^2(2) = 10.28$ ,  $p = 0.006$ ), reflecting varying differences in the strength of within-category and between-category priming, depending on the target category. Importantly, there was also a significant interaction of prime type by category match, reflecting greater priming within- than between-category ( $\chi^2(1) = 763.92$ ,  $p < 0.001$ ), as well as interactions of prime type by target category ( $\chi^2(2) = 22.72$ ,  $p < 0.001$ ), and category match and target category ( $\chi^2(2) = 48.12$ ,  $p < 0.001$ ). Unpacking these interactions, we looked at priming within-category and between-category separately.

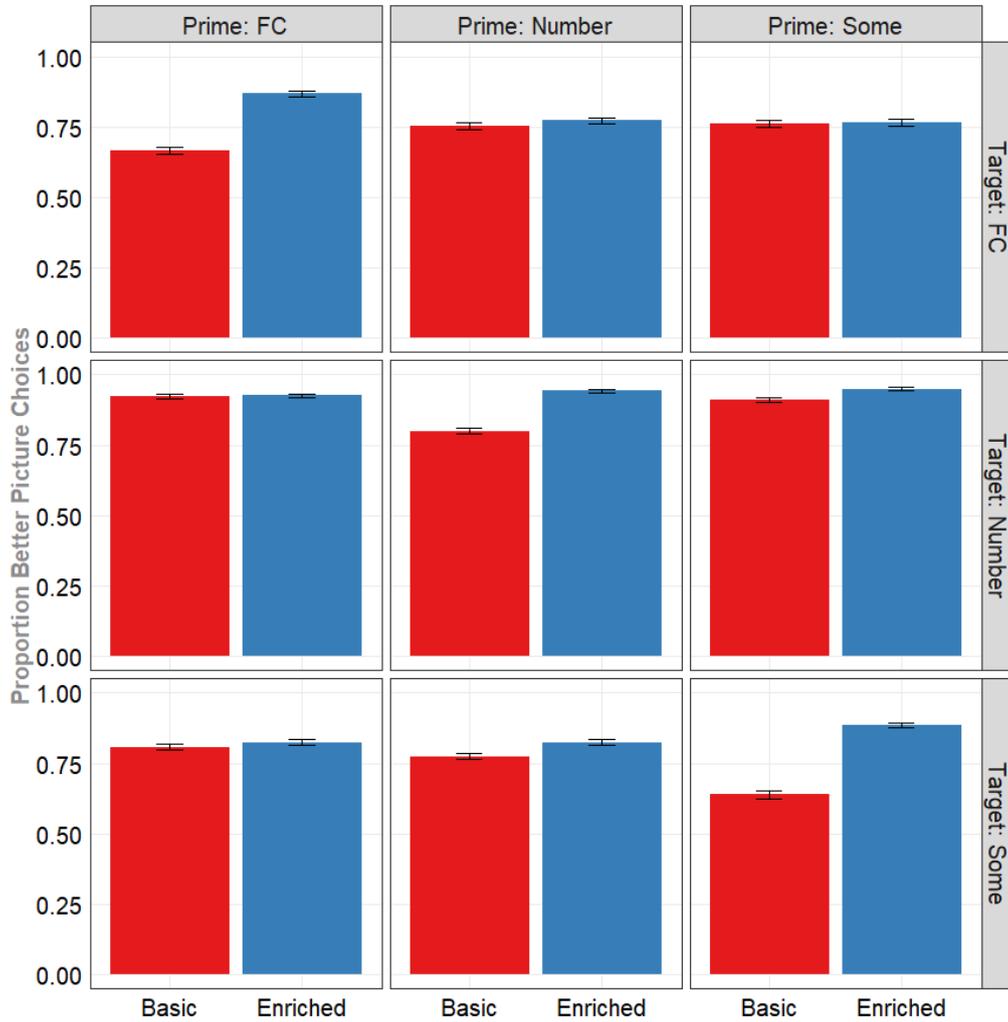


Figure 4: Results for the priming effects in Experiment 1. Bars show mean proportions of 'Better Picture' choices on target items after either enriched or basic primes. Vertical facets indicate the prime sentence category. Horizontal facets indicate the target sentence category. Error bars show  $\pm 1$  S.E., averaged by subjects.

#### 2.4.1. Within-Category Priming Effects

Including data from all three categories, a model with the fixed effects of prime type, category, and their interaction, showed highly significant main effects of both prime type ( $\chi^2(1) = 132.14$ ,  $p < 0.001$ ) and category ( $\chi^2(2) = 32.33$ ,  $p < 0.001$ ), but no significant interaction between them ( $\chi^2(2) = 5.13$ ,  $p = 0.077$ ). Simple effects analyses using treatment coding revealed a significant effect of prime type within each category separately (*some*:  $Z = -12.44$ ,  $p < 0.001$ ; *number*:  $Z = -9.76$ ,  $p < 0.001$ ; *FC*:  $Z = -9.50$ ,

$p < 0.001$ ). In sum, we found priming of enriched interpretations in all cases where the same sentence category appeared in both prime and target position, i.e., within-category, with no significant differences in the size of the effect between the three categories.

#### 2.4.2. Between-Category Priming Effects

To look at whether a common mechanism is responsible for generating the enriched reading of all three sentence categories, we looked at whether there were significant priming effects between categories. We initially combined all target categories, restricting the analysis just to between-category trials; for example, in cases where FC was the target category, we included only trials where *some* and number sentences were in prime position. We then built a model including prime type (basic vs. enriched), target category, and their interaction. This model revealed a significant main effect of category ( $\chi^2(2) = 60.67, p < 0.001$ ), a significant main effect of prime type ( $\chi^2(1) = 11.71, p < 0.001$ ), and a significant interaction of prime type with category ( $\chi^2(2) = 6.87, p = 0.032$ ). Note that both the interaction and the main effect of prime type could be the result of priming effects in one out of the three combinations of categories (e.g., priming between *some* and number, and not between either of these and FC sentences).

Because we are mainly interested in differences in between-category priming across different pairs of categories, we further broke down the results by each target category and conducted three separate analyses, looking again only at between-category trials. Each of these three models included fixed effects of both prime type and prime category, as well as their interaction. Looking at between-category trials with *some* targets, we found a significant interaction ( $\chi^2(1) = 7.49, p = 0.006$ ) and a significant main effect of prime type ( $\chi^2(1) = 6.44, p = 0.011$ ), reflecting greater priming from number primes to *some* targets than from FC primes to *some* targets. Simple effects analyses revealed a significant priming effect from number primes to *some* targets ( $Z = -3.53, p < 0.001$ ), but not from FC primes to *some* targets ( $Z = -0.97, p = 0.330$ ). A model of between-category trials with number targets found no significant main effects, but again, a significant interaction in the same direction ( $\chi^2(1) = 34.21, p < 0.001$ ). Simple effects analyses revealed a significant priming effect from *some* primes to number targets ( $Z = -5.19, p < 0.001$ ), but not from FC primes to number targets ( $Z = -0.24, p = 0.814$ ). Finally, looking at the between-category trials where FC sentences were the target, we found no significant main effects of either prime type or category, and a marginal interaction ( $\chi^2(1) = 3.08, p = 0.079$ ). Simple effects analyses did not reveal significant priming effects either from number to FC sentences ( $Z = -1.24, p = 0.216$ ) or from *some* to FC sentences ( $Z = 0.57, p = 0.568$ ). In sum, we found significant priming effects between number and *some* sentences in both directions, consistent with a shared computation underlying the enriched readings of both categories. In contrast, we found no between-category priming effects involving FC sentences, in any direction.

#### 2.5. Discussion

Experiment 1 shows evidence that some shared computation underlies the enriched readings of *some* and number words, as eliciting the enriched reading of one kind of sentence made participants more likely to derive the enriched reading of the other kind. At the same time, it also reveals significant differences in the relative accessibility of the enriched and basic readings between these two cases, suggesting some additional

difference in how the readings are generated in each case. In contrast, although an enriched reading of an FC sentence can prime the same reading of a subsequent FC sentence, we found no evidence of priming between FC sentences on the one hand, and sentences containing either *some* or number words on the other, consistent with there being no shared computation between the FC reading and the enriched readings in the other two cases.

### 3. Experiment 2

Experiment 1 showed that given the choice between the enriched and the basic interpretation of a sentence on enriched prime trials, participants preferred the enriched reading across all three categories. This raises the possibility that the basic primes, which forced participants to choose between a strongly dis-preferred basic picture and an outright false one, might be priming a general permissiveness towards semantic slop (as when a teacher with 19 or 21 students in their class might say that they have 20 students), or an increased tolerance for dis-preferred readings more broadly, or else an adaptation to the peculiarities of a specific speaker (Yildirim et al., 2016).

All of these alternative explanations predict that basic prime items would prime participants to accept an almost-true, but strictly speaking false interpretation of a sentence. In Experiment 2, we altered the number of entities shown in number target items, so that these items would present a choice between an almost-true picture and the *Better Picture* option. Recall that the number target items in Experiment 1 used the sentence *There are four Xs* paired with a picture showing 6 such Xs, and the *Better Picture* option. In Experiment 2, the control targets also used the sentence *There are four Xs* but paired it with a picture showing 3 Xs with the *Better Picture* option. If participants in Experiment 1 had been primed to accept a dis-preferred reading more generally, they should be at least as likely to accept this almost-true, but strictly false picture.

#### 3.1. Methods

*Participants.* We recruited an additional 400 participants on MTurk. 9 participants were removed because they indicated English was not their native language, and an additional 41 were removed for failing >10% of filler trials. 88 individual trials were excluded because a picture choice was made within 500 msec, less than 1% of all trials.

*Materials and Procedure.* The materials and procedure were identical to the number target condition of Experiment 1, including all three types of sentences and trials in prime position. As explained above, the only change was the number of objects in the basic picture on target items, from 6 to 3.

### 3.2. Results

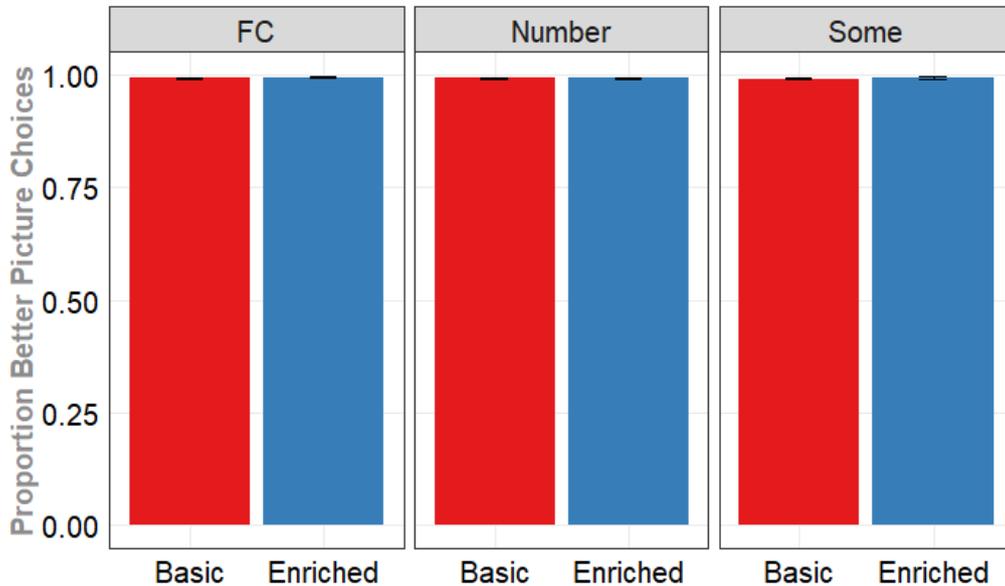


Figure 5: Results for the priming effects in Experiment 2. Bars show the proportion of *Better Picture* choices on target items, following either preceding enriched or basic primes within each sentence category. Error bars show  $\pm 1$  S.E., averaged by subjects.

Figure 5 shows participants' responses on target items. Participants rejected the picture with 3 Xs as a match to the sentence *There are four Xs*, and chose the *Better Picture* option, over 99% of the time in each case. A model of prime type (basic vs. enriched) and category match (between vs. within) showed no significant effects.

### 3.3. Discussion

Experiment 2 provides evidence that what was primed in Experiment 1 was not the shift between a sloppy and strict interpretation. In conjunction with the results from Experiment 1, we have strong evidence that the between- and within-category effects are due instead to a shared interpretative mechanism.

## 4. Experiment 3

In Experiment 1, we found no between-category priming involving FC sentences, together with significant priming between *some* and number sentences. A deflationary interpretation is that this priming pattern reflects patterns of picture similarity. The pictures for *some* and number items involve many of the same entities (houses, dogs, ponds, etc.), while the FC items show a figure, two roads, and black boxes, which the participant is told are covered trucks. Moreover, the basic pictures for both *some* and

number items always show more of the relevant entities than their counterpart enriched pictures, while this is not true for FC items. This alone might be enough to support a comparatively small between-category priming effect between *some* and number, making participants more likely to choose a basic picture on a target after having chosen a more visually similar basic picture on the prime. The results of Experiment 2 and of the Supplementary Experiment show that picture priming alone cannot be responsible for these effects, since no priming either within-, or between-category occurs if the picture stays the same but the accompanying sentences do not have an enriched reading (Suppl. Experiment), or when aspects of the picture change that are only relevant to the interpretation of the accompanying sentence, but not to the patterns of picture similarity across categories (Experiment 2). Nevertheless, it is possible that picture similarity is necessary, even if not alone sufficient, for between-category priming to occur. In that case, the lack of priming between FC sentences and either of the two categories could be due to the FC pictures being too different from both of the other categories.

To test this possibility, and to replicate Experiment 1, we created a new set of stimuli in which the very same picture matches one sentence from each of the three sentence categories on a given reading. For example, one picture could be used to depict the enriched reading of one *some* sentence, a different number sentence, and yet another FC sentence (in its FC reading). The same held for pictures corresponding to the basic reading (i.e. basic pictures) and pictures that are incompatible with either reading (i.e. false pictures). As a consequence, a given trial displayed all the same pictures regardless of whether it involved *some*, number, or FC sentences, and also regardless of whether it was part of a within- or between-category trial. If Experiment 3 replicates the results of Experiment 1, finding between-category priming between *some* and number, but not between either of these and FC, it would provide further evidence that priming between *some* and number is indeed due to a shared enrichment mechanism, and that the lack of between-category priming involving FC sentences cannot be due to differences in the degree of visual similarity between stimuli. It would also show that this pattern of priming is robust to changes in the specific stimuli.

*Participants.* We recruited 1200 new participants (400 per condition) using a combination of the Amazon Mechanical Turk and Testable Minds platforms. 22 participants were removed because they indicated English was not their native language, and an additional 391 participants were removed for failing >10% of filler trials (which were notably harder in this experiment). An additional 306 individual trials were excluded because a picture choice was made within 500 msec, less than 1% of trials.

*Materials and Procedure.* The logic of the trials and of the procedure was identical to Experiment 1, but all of the pictures were different. Every picture showed a figure on one of two floors of a building (upper or lower), with no access to the other floor. Each floor had the same two kinds of furniture (e.g., a table and a cabinet). Across the two floors, there were 4 objects of the same kind: bottles, umbrellas, vases, or hammers. However, one part of each object was uniquely colored (red, blue, yellow, or green), while the rest was black, showing the object's identity (e.g., the handle of each hammer was colored, but the characteristic head was black). In each picture, some of these objects were partly covered by being placed in an opaque bucket so that the common black part of the object was still visible, while the unique colored part was hidden. This allowed us to introduce

uncertainty for the FC sentences, just like covering trucks in Experiment 1 did, while keeping the kind of object identifiable.

Figure 6 shows an example of each kind of picture. For instance, the basic picture corresponds to the uncertainty/basic reading of an FC sentence, a number sentence, and a *some* sentence. This is achieved by using different components of the picture to verify each sentence. The *some* sentences concern which piece of furniture the objects are on, the number sentence concern the floor they are on, and the FC sentences concern which color hammers are visible vs. covered, and whether they are on the same floor as the figure.

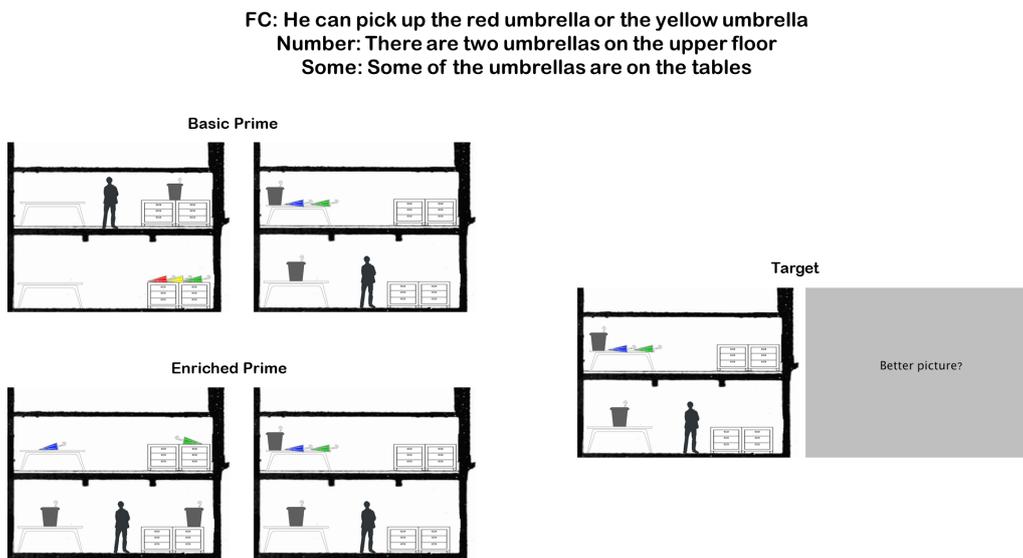


Figure 6: Sample prime and target items in Experiment 3. Participants saw either a basic prime or an enriched prime item, followed by a target item. Example sentences correspond to these items from each category. The same sentence can be paired with either type of prime or the target item. In basic prime items, the false picture is shown on the left and the basic picture is shown on the right. In enriched prime items, the enriched picture is shown on the left and the basic picture is shown on the right. The actual side of presentation of each picture on prime items was counterbalanced. On target items, the left-hand picture corresponds to the basic reading and the right-hand picture shows the *Better Picture* option. The side of presentation was fixed for all target items.

#### 4.1. Results

We analyze Experiment 3 using the same approach as in Experiment 1. We first ask whether participants were able to access both the basic and the enriched reading of sentences in each of the three categories. We then look for both within- and between-category priming effects.

*Accessibility of Basic and Enriched readings.* Figure 7 shows just the results from basic and enriched primes for each of the three categories, again reflecting the accessibility of the basic vs. enriched reading for each category. As in the corresponding Figure 3 from Experiment 1, ‘correct’ refers to the choice of the basic (rather than the false) picture on

basic primes, and the choice of the enriched (rather than the basic) picture on enriched primes. As figure 7 shows, participants were able to access both types of readings on their respective prime trials within each sentence category.

To test whether the accessibility of each reading differed between categories, we analyze the effect of category separately within enriched and basic primes, with the mediate level of each category coded as the reference level to enable comparisons both up and down. As in Experiment 1, these analyses showed that participants found the enriched reading of number sentences easiest to access, preferring this reading more strongly than in FC sentences ( $Z = 3.09$ ,  $p = 0.002$ ), and again for FC sentences more strongly than in the case of *some* sentences ( $Z = -9.21$ ,  $p < 0.001$ ). However, participants were more willing to choose the basic reading of number sentences on basic prime items. Indeed, they had a stronger preference for basic readings of number sentences than FC sentences ( $Z = 4.10$ ,  $p < 0.001$ ), and of FC sentences still more than *some* sentences ( $Z = 5.60$ ,  $p < 0.001$ ).

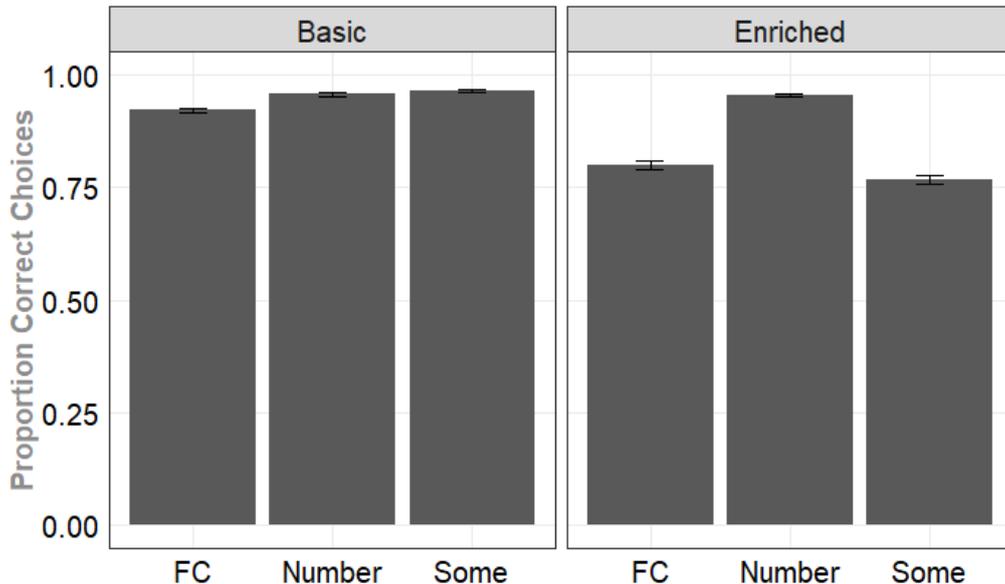


Figure 7: Responses on prime items in Experiment 3. Mean proportion of picture choices corresponding to the type of prime (basic or enriched). Error bars indicate +/- 1 standard error.

We further compared participants' behavior across experiments to test whether their stronger preference for enriched pictures on enriched number primes was less susceptible to changes in the stimuli than their choices on enriched *some* primes, consistent with prior findings that the enriched readings of number words are more strongly preferred. Combining data from both experiments, we selected just number and *some* prime items and analyzed the fixed effects of prime category (*some* vs. number) and Experiment (1 vs. 3). This analysis revealed a highly significant effect of prime category, reflecting more

choices of the enriched picture on number primes ( $\chi^2(2) = 787.64, p < 0.001$ ), a highly significant effect of Experiment, reflecting overall lower rates of enriched picture choices in Experiment 3 ( $\chi^2(1) = 503.42, p < 0.001$ ). Importantly, it also revealed a highly significant interaction, indicating a greater difference between *some* and number primes in Experiment 3 than in Experiment 1 ( $\chi^2(2) = 58.46, p < 0.001$ ), corresponding to a relatively more stable high rate of enriched picture choices on number primes in both experiments.

To get another estimate of the relative rates of enrichment between these two categories, we also looked at average responses on target items, independent of the effect of the preceding prime. We analyzed the effect of target category on the rate of *Better Picture* choices, comparing just *some* and number target items and finding a significantly higher rate of *Better Picture* choices for number ( $\chi^2(1) = 50.69, p < 0.001$ ). Finally, to compare the rate of *Better Picture* choices on target items across experiments, we combine data from both experiments and construct a new model including the effects of category (*some* vs. number), Experiment (1 vs. 3), and their interaction. We find significant fixed effects of target category ( $\chi^2(2) = 207.99, p < 0.001$ ), reflecting generally higher rates of *Better Picture* choices on number targets, of Experiment ( $\chi^2(1) = 385.39, p < 0.001$ ), reflecting an overall decrease in the rate of *Better Picture* choices in Experiment 3, and a significant interaction ( $\chi^2(2) = 17.81, p < 0.001$ ), reflecting a relatively smaller decrease in *Better Picture* choices for number compared to *some*. On targets as on enriched primes, participants were again reluctant to choose pictures corresponding to the *at least* readings of number sentences, and this tendency was relatively stable across experiments.

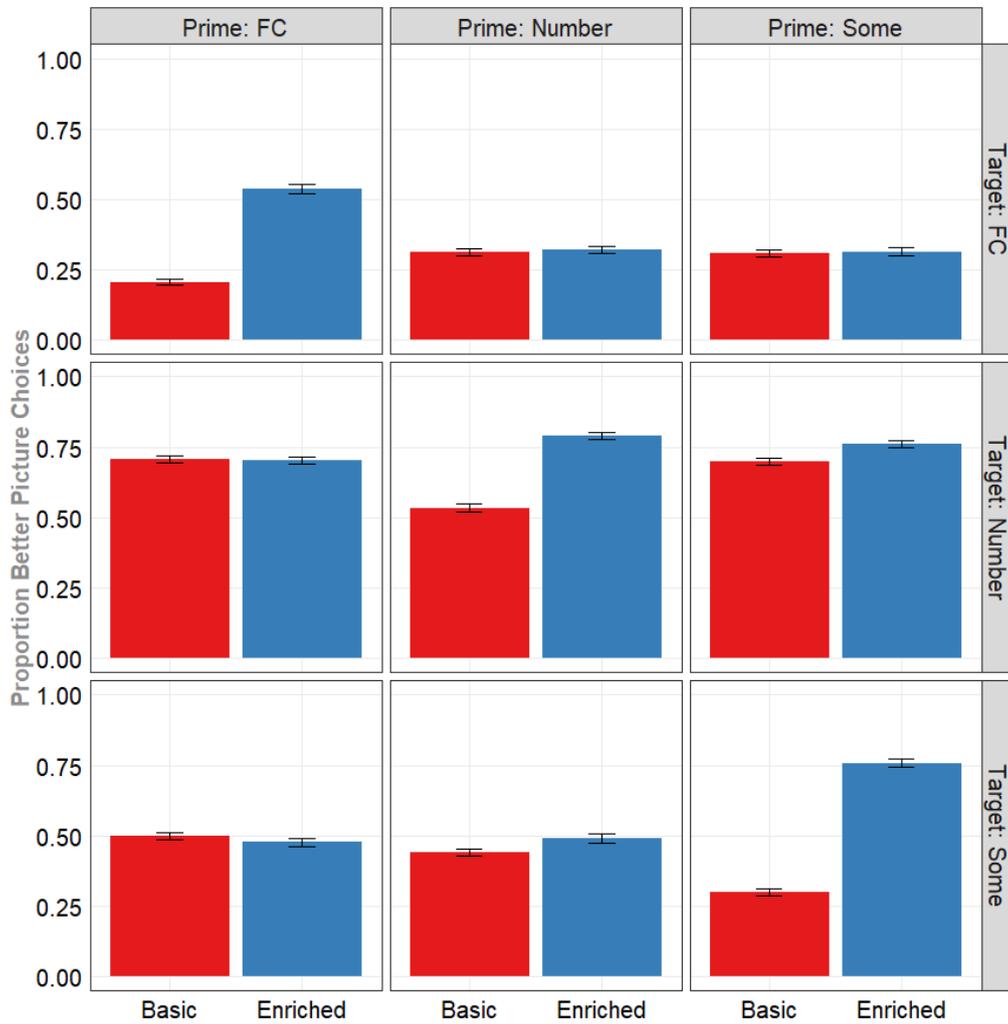


Figure 8: Results for the priming effects in Experiment 3. Bars show mean proportions of *Better Picture* choices on target items after either enriched or basic primes. Columns indicate the prime sentence category, rows indicate the target sentence category. Error bars show +/- 1 S.E., averaged by subjects.

*Priming.* Figure 8 shows the results of the priming manipulation, broken down by prime and target categories. To investigate priming effects both within- and between-category, we again first fit a model with fixed effects of target category (*some*, *number*, *FC*), category match (within vs. between) prime type (basic vs. enriched), and all interactions. We found that each interaction was highly significant; the three-way interaction ( $\chi^2(2) = 20.74, p < 0.001$ ), reflects varying differences in the strength of within-category and between-category priming, depending on the target category; the interaction of prime type by category match ( $\chi^2(1) = 829.96, p < 0.001$ ) reflects greater priming within- than

between-category. Unpacking these interactions, we looked at priming within-category and between-category separately, as in Experiment 1.

#### 4.1.1. Within-Category Priming

We again looked at whether there was a common within-category priming effect across all three sentence categories, and whether this effect differed between them. The pattern of results we find replicates Experiment 1. Looking only at within-category trials between subjects and in a model with the fixed effects of prime type, category, and their interaction, we find a highly significant main effect of prime type ( $\chi^2(1) = 468.88$ ,  $p < 0.001$ ), indicating higher rates of *Better Picture* choices following enriched primes compared to basic primes; a highly significant effect of category ( $\chi^2(2) = 123.11$ ,  $p < 0.001$ ), indicating the highest rate of *Better Picture* choices for number targets, and a significant interaction ( $\chi^2(2) = 9.82$ ,  $p = 0.007$ ), indicating differences in the size of the within-category priming effects between categories. Simple effects analyses once again revealed a highly significant effect of prime type in each category (some:  $Z = -15.64$ ,  $p < 0.001$ ; number:  $Z = -13.25$ ,  $p < 0.001$ ; FC:  $Z = -13.22$ ,  $p < 0.001$ ).

#### 4.1.2. Between-Category Priming

Here, too, we took the same approach as in Experiment 1, and again replicated the same pattern of results. Selecting only between-category trials, we analyzed the fixed effects of target category (*some*, number, FC), prime type (basic vs. enriched), and their interaction. This model revealed a significant main effect of target category ( $\chi^2(2) = 178.72$ ,  $p < 0.001$ ), a significant main effect of prime type ( $\chi^2(1) = 10.18$ ,  $p = 0.001$ ), and a marginal interaction of prime type with target category ( $\chi^2(2) = 5.61$ ,  $p = 0.060$ ).

Unpacking these interactions, we looked at between-category priming in each target category separately, with three identical models that included the fixed effects of prime category, prime type, and their interaction. We find a significant interaction between prime category and prime type in number targets ( $\chi^2(1) = 6.90$ ,  $p = 0.009$ ). Analysis of simple effects revealed a significant effect of prime type from *some* primes to number targets ( $Z = -2.73$ ,  $p = 0.006$ ), but no effect from FC primes to number targets ( $Z = -1.17$ ,  $p = 0.242$ ). Looking at between-category *some* target trials similarly revealed a significant interaction of prime type and prime category ( $\chi^2(1) = 8.63$ ,  $p = 0.003$ ). Again, simple effects analyses revealed a significant effect of prime type from number primes to *some* targets ( $Z = -2.41$ ,  $p = 0.016$ ), but not from FC primes to *some* targets ( $Z = 0.57$ ,  $p = 0.566$ ). Finally, looking at FC targets, we find neither a main effect of prime type ( $\chi^2(1) = 0.76$ ,  $p = 0.382$ ), nor an interaction between prime type and prime category ( $\chi^2(1) = 0.07$ ,  $p = 0.787$ ). Simple effects analyses confirmed that there was no significant effect of prime type either from number primes to FC targets ( $Z = -0.64$ ,  $p = 0.520$ ) or from *some* primes to FC targets ( $Z = -0.89$ ,  $p = 0.376$ ). In sum, we find the same pattern of results in Experiment 3 as in Experiment 1: significant priming effects between *some* and number sentences going in both directions, and no priming between either of these and FC sentences.

## 5. General Discussion

### 5.1. Summary of Results

We set out to test the hypothesis that enriched readings of *some* and number, together with FC readings, are derived through the computation of an implicature using a shared, category-general mechanism. We observed between-category priming effects from *some* to number and vice versa, such that accessing the enriched reading for one category increased the likelihood of accessing the enriched reading for the other category. We also found evidence that these effects are not reducible to either picture similarity or variable interpretive tolerance. Our results provide strong support for the view that the *exactly*-reading of number is derived through a mechanism that involves at least one sub-computation also used to derive the enriched reading of *some*, supporting implicature-based accounts of both categories.

However, our results also point to important differences between the two categories. Setting aside the priming effect, the preference for the enriched reading of number is stronger than the preference for the enriched reading of *some* relative to their respective basic readings (based on responses on enriched primes). Independently, the basic reading of number is more strongly dis-preferred than the basic reading of *some*. This is measured by the choice between a picture illustrating the basic readings vs. a picture that would falsify the relevant sentence under either reading (as in basic primes), and by the choice between a ‘Better Picture’ option vs. a picture illustrating the basic reading (as in target items, independently of the preceding prime). Lastly, while the overall rates of enrichment dropped from Experiment 1 to Experiment 3, which involved a relatively more difficult task, the decrease in enrichment rates was much more pronounced for *some* when compared to number (in line with Marty et al. 2013; De Neys & Schaeken 2007). Together, these results require a more nuanced analysis than simply accepting or rejecting a common implicature-based account for both *some* and number, as we will discuss below.

For FC sentences, we found that the choice between the FC reading and the uncertainty reading can be primed within-category, but not between-category by either *some* or number. Enriched FC primes likewise did not prime the enriched readings of either *some* or numbers. These results pose a challenge for accounts on which FC readings involve recursive application of the same implicature mechanism. As we discuss below, the results are consistent both with non-implicature based accounts of FC, and also with a newer implicature-based account which does not involve the negation of alternatives as a sub-computation.

Finally, this pattern of results rules out common deflationary explanations for structural priming effects. Experiment 3 rules out priming based on patterns of picture similarity, since a single picture matched the same reading of all three sentence categories, while there was only priming between *some* and number, and not FC sentences (see also the control experiment in Supplementary Materials). That the same priming pattern obtains in Experiment 1, in which the pictures do vary across sentence categories, also suggests that picture similarity is unlikely to be responsible for the results of that experiment. Finding the same pattern of results across two experiments using different materials also makes it unlikely that the results can be explained by differential priming of more similar verification strategies between a given reading of *some* and number than between either of these and the same reading of FC sentences. Such an explanation

would have to specify what verification strategies could be shared by the primes of *some* trials and the targets of number trials (and vice versa), but not between either of these and FC trials; and the explanation must then hold for both Experiments 1 and 3, or else explain why they coincidentally happen to produce the same complex pattern of results. We cannot see such an explanation.

### 5.2. The Case of *Some* vs. *Number*

Our results support the view that speakers use implicature computations when accessing the enriched readings of both *some* and number words. But how can this conclusion be reconciled with the differences between the two categories found in the current study, as well as in previous studies?

We propose that instead of viewing implicature as a single mechanism that is either used or not, it is best to consider the mechanism in terms of its two major sub-computations, which are in part independent. Viewed this way, there is a spectrum of related types of implicature. As discussed in the introduction, these two sub-computations are:

- (10) Core Implicature Mechanism
1. Generate alternative(s)
  2. Negate alternative(s)

The dimension along which the implicature spectrum extends is defined by whether each of these steps is *computed online*, or else is pre-computed, with its output stored. On one end of this spectrum, both the generation and negation of the alternatives are computed online every time. On the other end, both the generation and negation of relevant alternatives are done offline. Consider the output of the implicature mechanism for a sentence like, *Some of the houses have a fence*, namely, *Some of the houses have a fence & not all of the houses have a fence*. It results in a meaning for the quantifier *some* which can be paraphrased as *some but not all*. This meaning could be stored as a second lexical entry for the quantifier, eliminating the need to compute the corresponding implicature online. The same applies to a sentence containing a number word, like, *There are four dogs*. Generating and negating the alternative *There are five dogs* derives a meaning of the whole sentence which corresponds to an *exactly four* interpretation of the number word itself, and this could likewise be stored as one of its lexical meanings instead of being computed through an implicature computation every time.

An intermediate possibility is that instead of storing what amounts to an additional lexical entry for items like *some* and *four*, only the alternatives, i.e., the output of the first sub-computation, are stored together with the basic meaning of a lexical item like *some*. These alternatives can then be retrieved when the lexical item is accessed. This is the assumption behind so-called Horn scales (e.g. Horn, 1972). The second sub-computation, i.e., the negation of these alternatives, would then still be computed online whenever the enriched meaning is accessed. Note that the converse is not possible – alternatives cannot be generated online, but pre-negated and stored offline, because negating an alternative offline would require that alternative to be available offline as well. This means that, in sum, there are three possible points on a spectrum from storage to computation that any given case could occupy: both the generation and negation of alternatives could be

computed online, both could be pre-computed and their outputs stored as an additional lexical entry, or the alternatives could be stored while their negation is computed online.

Our results suggest that number words and *some* occupy different positions on this spectrum. The between-category priming effects we found for *some* and number provide strong evidence that, despite any efficiency that lexical storage of the enriched meanings might provide, the two enriched meanings do involve at least one shared online computation. At the same time, the enriched reading is easier to access for number words than for *some*, while the basic reading is harder to access, pointing to an important difference.

To explain both the similarity and the difference, we hypothesize that the alternatives of number words are stored offline, to be negated online when an implicature is called for, while in the case of *some*, both the generation and the negation of alternatives happens online. On this view, the computation that is shared between *some* and number must be the negation of alternatives, since only this step is computed online in both cases. This shared computation of negating alternatives is what gives rise to the between-category priming effect.<sup>5</sup> This view also explains why the enriched reading of number words is more persistent and strongly preferred than the enriched reading of *some*: the alternatives of number words are retrieved (but not always negated) obligatorily when the lexical meaning of a given number is accessed.

There is independent evidence that the alternatives for each number word are stored together with its basic meaning. Children first learn number words in a fixed, ascending order by rote memorizing the count list like a nursery rhyme, creating associations between the sounds of different number words before they know any of their meanings (Wynn, 1990; Carey, 2009). When they later begin to learn numerical meanings, they start with intuitions about how the meanings of unknown number words relate to each other (Feiman et al., 2019), and rely on the meanings of smaller number words that they learn first to understand the meanings of words higher in the count list (Barner, 2017; Barner & Bachrach, 2010). Thus, on multiple accounts, access to numerical alternatives is necessary to construct numerical meanings. Likewise, evidence from adults shows that they access numerical alternatives rapidly and automatically during comprehension, as in SNARC and numerical distance effects (Fias et al. 1996; Dehaene et al. 1990, for review see Dehaene 2011).

Moreover, in rapid estimation tasks adults do not consistently associate a single quantity (e.g., of dots) with a single number word. Rather, which number word they give as an estimate of one quantity is influenced by what other number word was previously given to describe another quantity (Sullivan & Barner, 2013; Izard & Dehaene, 2008). This suggests that mapping a given quantity to a number word affects mappings for other number words as well, consistent with the idea that the meaning of number words involves access to alternatives (for review, see Marchand & Barner 2018).

By contrast, the meaning of the existential quantifier *some* can be understood without

---

<sup>5</sup>Rees & Bott (2018) find that once the alternatives themselves are primed, there is no additional priming effect of negating those alternatives. They interpret these findings as evidence that the negation of alternatives cannot be primed, in apparent contrast to our proposal here. However, the separable effects of priming alternatives and priming negating them need not be cumulative; for example, there may be a low ceiling on how much can be primed by a combination of different factors. But crucially, since *some* and number use different alternatives, the between-category priming effect we observe is most consistent with our assumption that it is the negation of alternatives which is being primed here (see also Bott & Chemla (2016)).

any reference to alternatives like *all*. The alternative needs to be accessed only when the enriched reading is at stake. It may therefore be more efficient to compute the alternative online whenever this enriched reading is accessed, rather than storing it together with the meaning of *some* and unnecessarily accessing it every time *some* is retrieved.

The hypothesis that the alternatives for number are automatically retrieved whenever a speaker accesses its basic reading can explain why the rate of enriched meanings is higher for number than for *some*. This is because generating (and thus activating) alternatives is not just a sub-computation of the implicature mechanism, but may itself be a trigger for the implicature to be computed (Chierchia, 2013; Husband & Ferreira, 2016). And conversely, if the alternatives are not automatically accessed for *some*, but need to be computed online, then an overall lower rate of enriched readings, and slower processing times for these readings when compared to the enriched readings of number, are expected (see our discussion in section 1.1; e.g. Marty et al. 2013; Bott et al. 2012; Huang & Snedeker 2011; Panizza et al. 2009; Huang & Snedeker 2009; De Neys & Schaeken 2007; Breheny et al. 2006; Bott & Noveck 2004; Noveck & Posada 2003, cf. Degen & Tanenhaus 2015; Grodner et al. 2010; Feeney et al. 2004).

This view also entails that the debate between the structural theory of alternatives (Katzir, 2007) on the one hand and the scale-based view on the other (Horn, 1972) needs to be reconsidered. Rather than two competing hypotheses about how alternatives are derived *in general*, both mechanisms might be instantiated by different phenomena on the implicature-spectrum (cf. de Carvalho et al., 2016).

### 5.3. The Case of Free Choice

Contrary to *some* and number, we find no evidence of a shared implicature computation with FC readings. This is consistent with the idea that FC readings arise through a mechanism that is specific to the interaction of modals with a disjunction in their scope (Aloni, 2007; Simons, 2005), or through a lexically stored, non-standard meaning of *or* (Geurts 2005; Zimmermann 2000; see Meyer 2020 for an overview). However, the latter accounts do not straightforwardly extend to the cases we considered in our experiments, as they are specific to FC sentences in which the modal operator is interpreted epistemically or deontically, whereas for our stimuli, the most plausible interpretation of the modal was circumstantial (Kratzer, 2012). The accounts of Aloni and Simons do not exhibit this limitation and are thus consistent with our results without requiring additional assumptions.

There is however one implicature-based account which is consistent with our results. Bar-Lev & Fox (2017) propose that in addition to the implicature mechanism in (5) above, which involves the generation and negation of alternatives, there is also a mechanism which involves the generation and *assertion* of alternatives, as shown below (the parts that differ from (3) are underlined):

- (11) Non-negating implicature mechanism for a FC sentence *can* [*A or B*]
1. Generate alternatives (*can A*, *can B*)
  2. Assert these alternatives

Consider the predictions of this account for between-category priming with *some* and number. We argued above that the shared computation which was driving the priming effect between *some* and number must be the negation of alternatives. If this is correct,

and if the generation of alternatives cannot be primed at all, then this would predict the absence of a priming effect with FC, consistent with what we observed. If, on the other hand, the generation of alternatives can be primed in principle, then we would expect to find between-category priming between *some* and FC, contrary to what we in fact observed.<sup>6</sup>

Could our results nevertheless be reconciled with a (recursive) implicature account of FC (Fox, 2007; Alonso-Ovalle, 2006)?<sup>7</sup> First, one would have to assume that the negation of alternatives is not a computation which can be primed at all. To see why this assumption is necessary, recall that to derive FC readings, these accounts assume a recursive implicature computation by which first the alternatives to the FC disjunction are generated, then their enriched reading is computed (by an additional implicature computation), and finally these enriched alternatives are negated (see section 1.2). Construing the enriched alternatives itself involves the negation of alternatives, as we have seen above. Deriving the FC reading will therefore involve the negation of alternatives – twice. Therefore, if negation of alternatives can be primed, we would expect to see between-category priming effects between *some*, number and FC, rather than just between the former two categories.

Assume then that negation cannot be primed. Can our results be reconciled with the (recursive) implicature account of FC under this assumption? The observed priming effects between *some* and number would then be due to the shared computation of generating or using alternatives. But as we have seen above, this computation is also involved in the FC reading under the implicature account, leaving the lack of priming effects between FC and the other two sentence categories unexplained yet again.

The simple hypothesis that only the generation and/or usage of alternatives was primed thus needs additional assumptions in order to be consistent with the implicature-based account of FC. Are there any plausible differences in the way alternatives are generated or used in *some*/number and FC? While we suggested above that the alternatives are stored together with the lexical meaning for number, but generated online for *some* when an enriched reading is accessed, it is at least conceivable that contrary to this hypothesis, the alternatives are lexically encoded (as hypothesized in the original Horn-scale approach) for both number and *some*, and thus not generated online at all in either case. By contrast, the alternatives of FC sentences *must* be generated online, depending on whatever lexical material the relevant FC sentence contains. This makes FC inferences a special case of so-called *ad hoc* implicatures (Hirschberg, 1991), in which the alternatives are generated using the immediate linguistic material in the relevant sentence itself.

Could this explain the lack of between-category priming with FC sentences? Previous results from Bott & Chemla (2016) weigh against this hypothesis. Bott & Chemla show that *ad hoc* implicatures can prime and be primed by enriched readings of *some* and number. This entails that either all of *some*, number, and *ad hoc*-implicatures involve

---

<sup>6</sup>Another recent suggestion is the hybrid account of Champollion et al. (2019), which derives FC readings through a combination of probabilistic/game-theoretic reasoning and implicature computation. By itself this would wrongly predict between-category priming with *some* and Number. However, the exact predictions depend on how the model would derive uncertainty readings, and are therefore difficult to assess.

<sup>7</sup>The simple, embedded implicature mechanism of Klindinst (2007) also involves the generation and negation of alternatives, which is why our argument in what follows applies to this account as well.

the online generation of alternatives after all, in which case we would expect priming with FC as well. Alternatively, it could mean that priming effects arise across categories even when one category involves lexically-stored alternatives and the other involves the online generation of alternatives. But this just means that we would again expect FC to exhibit priming with *some* and number, contrary to fact.<sup>8</sup>

We think that a more plausible explanation that is consistent with a recursive implicature account is the following. Recall that our FC basic primes presented a choice between a basic/uncertainty picture in which only one disjunct was true (though it was not revealed which one), and a false picture in which none were true. Going back to our example item from Experiment 3, for the sentence *He can pick up the red vase or the yellow vase*, one picture presented the person as being able to pick up one of the relevant vases, whereas in the other picture he could not pick up any of the vases. In the FC enriched primes, the choice was between an enriched/FC picture in which the person could pick up both relevant vases, and a basic/uncertainty picture in which only one relevant vase was reachable. We did not provide any information as to whether it was possible to pick up two vases. Therefore, while the basic picture always also verified the additional exclusivity implicature *It is not the case that he can pick up the red vase and the yellow vase*, the FC picture was under-determined with respect to this implicature. Consider now how this could have influenced the between-category priming pattern. Assume that participants had indeed derived the additional exclusivity implicature in basic FC primes, but not in enriched FC primes. If the FC reading was derived through a separate, non-implicature mechanism, we would have expected to observe an *anti-priming* effect between FC and both *some* and number, as only the basic prime trials would involve computing an exclusivity implicature. If, on the other hand, the FC reading was derived through a recursive implicature mechanism, the expected priming effect could have been masked, as the basic, uncertainty primes would have contained a (simple) computation of the exclusivity implicature, while the enriched, FC primes would have contained a recursive implicature computation to derive the FC reading. The differences between a simple and a recursive implicature computation might not be pronounced enough, so that both types of FC primes would have primed implicature computations, resulting in a lack of priming effect overall. Still, this does require assuming that readings that involve a recursive application of implicature are neither a better prime nor a more sensitive target than readings that rely on just one implicature computation. Future research could investigate whether both basic and enriched FC primes increase the rate of enriched readings of number and *some* targets relative to a baseline in which the primes are replaced with filler items that involve no implicature computations.

While we cannot completely rule out an implicature-based account of FC readings, our results are more clearly consistent with two other types of accounts of FC: those which posit interpretative mechanisms specific to the interaction between a modal operator and a disjunction, and a novel implicature-based account which assumed the *assertion* rather than the negation of alternatives as a sub-step of the implicature mechanism.

---

<sup>8</sup>Parallel reasoning applies to our discussion of Bar-Lev & Fox (2017) above. We argued there that if the generation of alternatives cannot be primed, then our results are consistent with their inclusive implicature account of FC. To maintain the account even under the assumption that generation of alternatives can be primed in principle, Bar-Lev & Fox would have to maintain that differences in the way alternatives are generated lead to a lack of priming, which would again conflict with the results from Bott & Chemla (2016).

#### 5.4. Further Methodological Contributions

##### 5.4.1. Picture Priming

Experiment 3 established that picture similarity appeared to play no role in either within- or between-category priming effects. With all trials containing the same pictures, we found the priming effect not merely reduced between FC and the other categories, but absent altogether, replicating our findings from Experiment 1. While picture similarity did not seem to play a role in these experiments, controls for picture similarity will continue to be necessary in each new task, at least until the factors that drive priming based on visual similarity in structural priming tasks are better understood. To this end, it may help to consider why we found no priming based on picture similarity here. We designed the stimuli in Experiment 3 so that each type of sentence would require verification from different aspects of the same picture (e.g., looking at how many objects are on which floor to verify number sentences, but looking at the proportion of objects on each type of furniture across both floors to verify sentences with *some*). We suspect that participants may not have been primed by picture similarity across sentence types because they were attending to different parts of the picture in each case. This suggests that semantic interpretation of sentences takes precedence over visual categorization when participants are asked to match a picture to a sentence, such that visual similarity cannot override interpretative choices. It further suggests that structural priming studies may need to worry less about sheer visual similarity between stimuli, and more about shared verification strategies relative to the linguistic materials (see also Feiman et al. 2020), a concern which the design of Experiment 3 also rules out in this case.

##### 5.4.2. Rates of Implicature with Different Stimuli

Our experiments add important insights into the role of stimuli in the overall rate at which participants compute implicatures, i.e., access enriched readings. We find a very high rate of enrichment in Experiment 1 compared to Bott & Chemla (2016), perhaps because we use naturalistic displays instead of configurations of abstract symbols. By contrast, the rate of enrichment is much lower in Experiment 3, even though we still use more naturalistic stimuli. This difference could be due to the complexity of the task, in line with other results indicating that cognitive load lowers the rate at which speakers compute implicatures (Marty et al., 2013; De Neys & Schaeken, 2007).

##### 5.4.3. Stimuli for Assessing Free Choice Readings

Our results demonstrate that the stimuli we developed for both Experiment 1 and Experiment 3 provide a useful way of investigating interpretations of FC disjunctions. Participants' performance on both kinds of primes indicates that they had no difficulties with these items, even though they involved a complex manipulation including covered objects and a partially ignorant narrator. This manipulation constitutes an essential improvement compared to other items used to elicit FC readings (e.g. Chemla & Bott 2014), as it made both the uncertainty and the FC reading pragmatically felicitous. On target trials, we obtained a relatively high rate of responses consistent with an uncertainty reading ( $\sim 75\%$ ), as well as a within-category priming effect, which suggests our manipulation was successful at making both readings available and felicitous.

### 5.5. Conclusion

Implicature is usually thought of as all-or-nothing – either a given meaning is derived through implicature, or it isn't. We have argued for a different perspective. The two main sub-computations in the implicature mechanism – generating alternatives and negating them – can be considered independently, leading to a view of implicature as a spectrum. On one end of this spectrum, both are computed online. On the other, the output of both computations is stored as a separate lexical entry. In the middle, only the negation, but not the generation of alternatives is computed online. This finer-grained perspective on implicature allows us to reconcile a seemingly conflicting pattern of results. Extending our investigation to Free Choice disjunctions, we find no evidence of any shared sub-computations between this and more canonical cases of implicatures such as *some* and *number*, constraining theories that explain Free Choice as a case of implicature.

### References

- Aloni, M. (2007). Free Choice, modals, and imperatives. *Natural Language Semantics*, 15, 65–94. doi:10.1007/s11050-007-9010-2.
- Alonso-Ovalle, L. (2006). *Disjunction in Alternative Semantics*. Doctoral dissertation UMass Amherst.
- Bar-Lev, M., & Fox, D. (2017). Universal free choice and innocent inclusion. In *Proceedings of SALT XXVII* (pp. 95–115).
- Barner, D. (2017). Language, procedures, and the non-perceptual origin of natural number concepts. *Journal of Child Language*, 44, 553–590.
- Barner, D., & Bachrach, A. (2010). Inference and exact numerical representation in early language development. *Cognitive Psychology*, 60, 40–62. doi:10.1016/j.cogpsych.2009.06.002.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67, 1–48. doi:10.18637/jss.v067.i01.
- Benz, A. (2006). Utility and relevance of answers. In A. Benz, G. Jaeger, & R. van Rooij (Eds.), *Game Theory and Pragmatics* (pp. 195–219). Hampshire: Palgrave.
- Bergen, L., Levy, R., & Goodman, N. (2016). Pragmatic reasoning through semantic inference. *Semantics and Pragmatics*, 9, 1–84. doi:10.3765/sp.9.20.
- Bott, L., Bailey, T., & Grodner, D. (2012). Distinguishing speed from accuracy in scalar implicatures. *Journal of Memory and Language*, 66, 123–142. doi:10.1016/j.jml.2011.09.005.
- Bott, L., & Chemla, E. (2016). Shared and distinct mechanisms in deriving linguistic enrichment. *Journal of Memory and Language*, 91, 117–140. doi:10.1016/j.jml.2016.04.004.
- Bott, L., & Noveck, I. (2004). Some utterances are underinformative: The onset and time course of scalar inferences. *Journal of Memory and Language*, 51, 437–457. doi:10.1016/j.jml.2004.05.006.
- Branigan, H., & Pickering, M. (2017). An experimental approach to linguistic representation. *Behavioral and Brain Sciences*, 40, E282. doi:10.1017/S0140525X16002028.
- Breheny, R., Ferguson, H., & Katsos, N. (2013). Investigating the timecourse of accessing conversational implicatures during incremental sentence interpretation. *Language and Cognitive Processes*, 28, 443–467. doi:10.1080/01690965.2011.649040.
- 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, 434–464. doi:10.1016/j.cognition.2005.07.003.
- Carey, S. (2009). *The Origin of Concepts*. New York: Oxford University Press.
- Carston, R. (1998). Informativeness, relevance and scalar implicature. In R. Carston, & S. Uchida (Eds.), *Relevance theory: Applications and implications* (pp. 179–236). Amsterdam: Benjamins.
- de Carvalho, A., Reboul, A., van der Henst, J.-B., Cheylus, A., & Nazir, T. (2016). Scalar implicatures: The psychological reality of scales. *Frontiers in Psychology*, 7, 1500. doi:10.3389/fpsyg.2016.01500.
- Chemla, E., & Bott, L. (2014). Processing inferences at the semantics/pragmatics frontier: Disjunctions and Free Choice inferences at the semantics/pragmatics frontier: Disjunctions and Free Choice. *Cognition*, 130, 380–396. doi:10.1016/j.cognition.2013.11.013.
- Chierchia, G. (2004). Scalar implicatures, polarity phenomena, and the syntax/pragmatics interface. In A. Belletti (Ed.), *Structures and Beyond* (pp. 39–103). Oxford University Press.
- Chierchia, G. (2013). *Logic in Grammar*. Oxford University Press.

- Chierchia, G., Crain, S., Guasti, M., Gualmini, A., & Meroni, L. (2001). The acquisition of disjunction: Evidence for a grammatical view of scalar implicatures. In A.-J. Do, L. Dominguez, & A. Johansen (Eds.), *Proceedings of the 25th Annual Boston University Conference on Language Development* (pp. 157–168). Somerville, MA: Cascadia Press.
- De Neys, W., & Schaeken, W. (2007). When People are More Logical Under Cognitive Load. *Experimental Psychology*, *54*, 128–133.
- Degen, J., & Tanenhaus, M. (2015). Processing scalar implicature: A constraint-based approach. *Cognitive Science*, *39*, 667–710. doi:10.1111/cogs.12171.
- Dehaene, S. (2011). *The Number Sense: How the Mind Creates Mathematics*. Oxford, UK: Oxford University Press.
- Dehaene, S., Dupoux, E., & Mehler, J. (1990). Is numerical comparison digital? analogical and symbolic effects in two-digit number comparison. *Journal of Experimental Psychology: Human Perception and Performance*, *16*, 626.
- Feeney, A., Scafton, S., Duckworth, A., & Handley, S. J. (2004). The story of some: Everyday pragmatic inferences by children and adults. *Canadian Journal of Experimental Psychology*, *58*, 121–132.
- Feiman, R., Hartshorne, J., & Barner, D. (2019). Contrast and entailment: Abstract logical relations constrain how 2- and 3-year-old children interpret unknown numbers. *Cognition*, *183*, 192–207.
- Feiman, R., Maldonado, M., & Snedeker, J. (2020). Priming quantifier scope: Reexamining the evidence against scope inversion. *Glossa*, *5*.
- Fias, W., Brysbaert, M., Geypens, F., & d’Ywalle, G. (1996). The importance of magnitude information in numerical processing: Evidence from the SNARC effect. *Mathematical Cognition*, *2*, 95–110.
- Foppolo, F., & Marelli, M. (2017). No delay for *some* inferences. *Journal of Semantics*, *34*, 659–681. doi:10.1093/jos/ffx013.
- Fox, D. (2007). Free Choice and the theory of scalar implicatures. In U. Sauerland, & P. Stateva (Eds.), *Presupposition and Implicature in Compositional Semantics* (pp. 71–120). New York: Palgrave Macmillan.
- Franke, M. (2011). Quantity implicatures, exhaustive interpretation, and rational conversation. *Semantics and Pragmatics*, *4*, 1–82. doi:10.3765/sp.4.1.
- Gazdar, G. (1979). *Pragmatics: Implicature, Presupposition and Logical Form*. New York: Academic Press.
- Geurts, B. (2005). Entertaining alternatives: Disjunctions as modals. *Natural Language Semantics*, *13*, 383–410. doi:10.1007/s11050-005-2052-4.
- Geurts, B. (2010). *Quantity Implicatures*. Cambridge University Press.
- Goodman, N., & Frank, M. (2016). Pragmatic language interpretation as probabilistic inference. *Trends in Cognitive Sciences*, *20*, 818–829.
- Grice, H. P. (1975). Logic and Conversation. In D. Davidson, & G. Harman (Eds.), *The Logic of Grammar* (pp. 64–75). Encino, CA: Dickenson.
- Grice, H. P. (1989). *Studies in the Way of Words*. Cambridge, MA: Harvard University Press.
- Grodner, D., Klein, N., Carbary, K., & Tanenhaus, M. (2010). “Some”, and possibly all, scalar inferences are not delayed: Evidence for immediate pragmatic enrichment. *Cognition*, *116*, 42–55. doi:10.1016/j.cognition.2010.03.014.
- Gualmini, A., Crain, S., Meroni, L., Chierchia, G., & Guasti, M. (2001). At the Semantics/Pragmatics Interface in Child Language. In R. Hastings, B. Jackson, & Z. Zvolensky (Eds.), *Proceedings of SALT XI* (pp. 231–247). Ithaca, NY: Cornell University.
- Guasti, M., Chierchia, G., Crain, S., Foppolo, F., Gualmini, A., & Meroni, L. (2005). Why children and adults sometimes (but not always) compute implicatures. *Language and Cognitive Processes*, *20*, 667–696.
- Hirschberg, J. (1991). *A Theory of Scalar Implicature*. New York: Garland.
- Horn, L. (1972). *On the Semantic Properties of Logical Operators in English*. Ph.D. thesis University of California, Los Angeles.
- Horn, L. (1989). *A Natural History of Negation*. Chicago, IL: University of Chicago Press.
- Huang, Y. T., & Snedeker, J. (2009). Online interpretation of scalar quantifiers: Insight into the semantics-pragmatics interface. *Cognitive Psychology*, *58*, 376–415. doi:10.1016/j.cogpsych.2008.09.001.
- Huang, Y. T., & Snedeker, J. (2018). Some inferences still take time: Prosody, predictability, and the speed of scalar implicatures. *Cognitive Psychology*, *102*, 105–126. doi:10.1016/j.cogpsych.2018.01.004.
- Huang, Y. T., Spelke, E., & Snedeker, J. (2013). What exactly do numbers mean? *Language Learning and Development*, *9*, 105–129. doi:10.1080/15475441.2012.658731.

- Huang, Y.-T. J., & Snedeker, J. (2011). Logic and conversation revisited: Evidence for a division between semantic and pragmatic content in real-time language comprehension. *Language and Cognitive Processes*, *26*, 1161–1172.
- Husband, M., & Ferreira, F. (2016). The role of selection in the comprehension of focus alternatives. *Language, Cognition and Neuroscience*, *31*, 217–235.
- Izard, V., & Dehaene, S. (2008). Calibrating the mental number line. *Cognition*, *106*, 1221–1247. doi:10.1016/j.cognition.2007.06.004.
- Katzir, R. (2007). Structurally-defined alternatives. *Linguistics and Philosophy*, *30*, 669–690. doi:10.1007/s10988-008-9029-y.
- Klinedinst, N. (2007). *Plurality and Possibility*. Ph.D. thesis University of California, Los Angeles.
- Kratzer, A. (2012). *Modals and Conditionals: New and Revised Perspectives*. Oxford: Oxford University Press.
- Kratzer, A., & Shimoyama, J. (2002). Indeterminate pronouns: The view from Japanese. In Y. Otsu (Ed.), *Third Tokyo Conference on Psycholinguistics* (pp. 1–25). Tokyo: Hituzi Syobo.
- van Kuppevelt, J. (1996). Inferring from topics: Scalar implicatures as topic-dependent inferences. *Linguistics and Philosophy*, *19*, 393–443. doi:10.1007/BF00630897.
- Mahowald, K., James, A., Futrell, R., & Gibson, E. (2016). A meta-analysis of syntactic priming in language production. *Journal of Memory and Language*, *in press*.
- Marchand, E., & Barner, D. (2018). Analogical mapping in numerical development. In D. Berch, D. Geary, & K. M. Koepke (Eds.), *Language and Culture in Mathematical Cognition* (pp. 31–47). Academic Press.
- Marty, P., Chemla, E., & Spector, B. (2013). Interpreting numerals and scalar items under memory load. *Lingua*, *133*, 152–163.
- Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing type I error and power in linear mixed models. *Journal of Memory and Language*, *94*, 305–315. doi:10.1016/j.jml.2017.01.001.
- Merin, A. (1999). Information, relevance, and social decisionmaking: Some principles and results of decision-theoretic semantics. In L. Moss, J. Ginzburg, & M. de Rijke (Eds.), *Logic, Language and Computation 16* (pp. 179–221). Stanford: CSLI Publications volume 2.
- Meyer, M.-C. (2013). *Ignorance and Grammar*. Doctoral dissertation MIT.
- Meyer, M.-C. (2020). Free Choice disjunction: *An Apple or a Pear*. In D. Gutzmann, L. Matthewson, C. Meier, H. Rullmann, & T. Zimmerman (Eds.), *The Wiley Blackwell Companion to Semantics*. Wiley.
- Noveck, I., Chierchia, G., Chevaux, F., Guelminger, A., & Sylvestre, E. (2002). Linguistic-pragmatic factors in interpreting disjunctions. *Thinking and Reasoning*, *8*, 297–326.
- Noveck, I., & Posada, A. (2003). Characterizing the time course of an implicature: An evoked potentials study. *Brain and Language*, *85*, 203–210.
- Panizza, D., Chierchia, G., & Clifton Jr., C. (2009). On the role of entailment patterns and scalar implicatures in the processing of numerals. *Journal of Memory and Language*, *61*, 503–518.
- Parikh, P. (2001). *The Use of Language*. Chicago: CSLI Publications.
- Potts, C., Lassiter, D., Levy, R., & Frank, M. (2015). Embedded implicatures as pragmatics inferences under compositional lexical uncertainty. Ms., Stanford University.
- R Core Team (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing Vienna, Austria.
- van Rooij, R., & Schulz, K. (2004). Exhaustive interpretation of complex sentences. *Journal of Logic, Language and Information*, *13*, 491–519.
- van Rooij, R., & Schulz, K. (2006). Pragmatic meaning and non-monotonic reasoning. the case of exhaustive interpretation. *Linguistics and Philosophy*, *29*, 205–250. doi:10.1007/s10988-005-3760-4.
- van Rooy, R. (2002). Relevance only. In J. Bos, M. Foster, & C. Matheson (Eds.), *Proceedings of the Sixth Workshop on the Semantics and Pragmatics of Dialogue* (pp. 155–160). Edinburgh.
- Sauerland, U. (2004). Scalar implicatures in complex sentences. *Linguistics and Philosophy*, *27*, 367–391. doi:10.1023/b:ling.0000023378.71748.db.
- Simons, M. (2005). Dividing things up: The semantics of ‘or’ and the modal/‘or’ interaction. *Natural Language Semantics*, *13*, 217–316. doi:10.1007/s11050-004-2900-7.
- Sperber, D., & Wilson, D. (1986). *Relevance: Communication and Cognition*. Cambridge, MA: Harvard University Press.
- Sullivan, J., & Barner, D. (2013). How are number words mapped to approximate magnitudes? *The Quarterly Journal of Experimental Psychology*, *66*, 389–402. doi:10.1080/17470218.2012.715655.
- van Tiel, B., van Miltenburg, E., Zevakhina, N., & Geurts, B. (2014). Scalar diversity. *Journal of*

- Semantics*, 0, 1–39. doi:10.1093/jos/ffu017.
- Tomlinson, J., Bailey, T., & Bott, L. (2013). Possibly all of that and then some: Scalar implicatures are understood in two steps. *Journal of Memory and Language*, 69, 18–35. doi:10.1016/j.jml.2013.02.003.
- Wynn, K. (1990). Children’s understanding of counting. *Cognition*, 36, 155–193. doi:10.1016/0010-0277(90)90003-3.
- Yildirim, I., Degen, J., Tanenhaus, M., & Jaeger, T. (2016). Talker-specificity and adaptation in quantifier interpretation. *Journal of Memory and Language*, 87, 128–143. doi:10.1016/j.jml.2015.08.003.
- Zimmermann, T. E. (2000). Free Choice disjunction and epistemic possibility. *Natural Language Semantics*, 8, 255–290. doi:10.1023/A:1011255819284.

## 6. Appendix 1: Practice Phase for Experiments 1-3

### 6.1. Practice Phase for Experiments 1 and 2

*NB: Every item (marked –) appeared on a separate screen. Square brackets describe accompanying pictures. "vs." denotes the presence of two picture options and a forced choice between them*

- Hi! Thanks for helping me out today! Click NEXT to start the experiment.
- I’m always going to tell you what I know. YOU have to guess which picture I’m talking about!
- Let’s practice!
- Three houses are green [3 green houses vs. 1 green house]
- You should choose the ‘Better Picture’ option if you feel the picture doesn’t capture the meaning of my sentence well enough.
- There are birds [scene with no birds vs. ‘Better Picture’]
- Now let’s meet my friend
- Sometimes, I know where he can go
- Here, he can go to the red truck or to the yellow truck [unlocked roads to both trucks]
- And here, he can go to the yellow truck but not to the red truck [unblocked road to yellow truck; blocked road to red truck]
- My friend always works with a red truck and a yellow truck. There are no other trucks
- But sometimes, I can’t see the trucks because they are covered, so I don’t know which is which. I will tell you all I know
- Now I can’t see the trucks. But I will tell you what I know: My friend is near the yellow truck or near the red truck [two covered trucks, person can only access one of them]
- Let’s see which one it was!
- Look, it was the yellow truck! [same picture with trucks uncovered; the truck person could access is yellow, inaccessible truck is red]
- Here is another picture
- My friend is near the yellow truck or near the red truck [two covered trucks, person can only access one of them]
- Let’s see which one it was!
- Look, it was the red truck! [same picture with trucks uncovered; the truck person could access is red, inaccessible truck is yellow]
- Now I will describe where my friend can go
- He can go to the red truck or to the yellow truck [two covered trucks, person can only access one of them]

- Here is another picture
- He can go to the yellow truck or to the red truck [two covered trucks, person can access both of them]
- Remember, I always tell you what I know, and YOU have to guess which picture I'm talking about!
- Let's practice one more time!
- He can go to the red truck but not to the yellow truck [both trucks accessible vs. red truck accessible and yellow truck inaccessible]
- Also remember that you should choose the 'Better Picture' option if you feel the picture doesn't capture the meaning of my sentence well enough.
- He can go to the yellow truck or to the red truck [two covered trucks, both inaccessible vs. 'Better Picture' option]
- Now you're ready to start!

### 6.2. Practice Phase for Experiment 3

*NB: Every item (marked -) appeared on a separate screen.*

- Hi! Thanks for helping me out today! Click NEXT to start the experiment
- This is my friend: [figure]
- He lives on one of the floors in this building: [entire building]
- He doesn't have access to the floors above or below him!
- I can see what's going on on all of the floors. Let's zoom in!
- The rooms have tables and cabinets [furniture]
- And there are some objects in the rooms
- Here, my friend can pick up the red hammer or the yellow hammer. He cannot get the green or the blue hammer because they are on the other floor [red, yellow hammer on floor with figure; blue, green hammer on other floor]
- And here, he can pick up the red umbrella, but not the yellow one [red umbrella on floor with figure; yellow umbrella on other floor]
- Did you notice the four colors? Red, yellow, blue and green. There is always one object in each color. There won't be any other colors [all objects in 4 colors]
- You see: There is a red hammer, a yellow hammer, a green one, and a blue one [objects mentioned in rooms]
- Sometimes, I can't see the color of the objects. But I will tell you all I know!
- Now, I can't see the color of all the vases, but I will tell you what I know: My friend can pick up the green vase or the blue vase [1 partially covered vase on floor with figure, red, yellow, and 1 partially covered vase on other floor]
- Lets see which one it is!
- Look, it's the blue vase! [same picture, vases now fully visible]
- He can pick up the green hammer or the blue hammer [1 partially covered hammer on floor with figure; red, yellow, and 1 partially covered hammer on other floor]
- Look, it's the green hammer! [now fully visible]
- He can pick up the red vase or the yellow vase [2 vases partially covered on floor with figure, blue, green hammer visible on other floor]
- Look, I was right! [now fully visible]
- Now let's get started. I am always going to tell you what I know. YOU have to guess which picture I'm talking about. Let's practice!
- There are three bottles on the upper floor [true vs. false picture]

- And you should choose the 'Better Picture' option if you feel the picture doesn't capture the meaning of my sentence well enough
- There are hammers [false picture vs. 'Better Picture']
- Remember, you have to guess which picture I'm talking about. Let's practice one more time!
- All of the umbrellas are on the tables [true vs. false picture]
- Now you're ready to start!

## 7. Appendix 2: Filler Items in Experiments 1-3

### 7.1. Number Fillers

#### Number Filler Type I

There are six houses



Figure 9: Exp 1 Number Filler Type I

There are three bottles on the upper floor

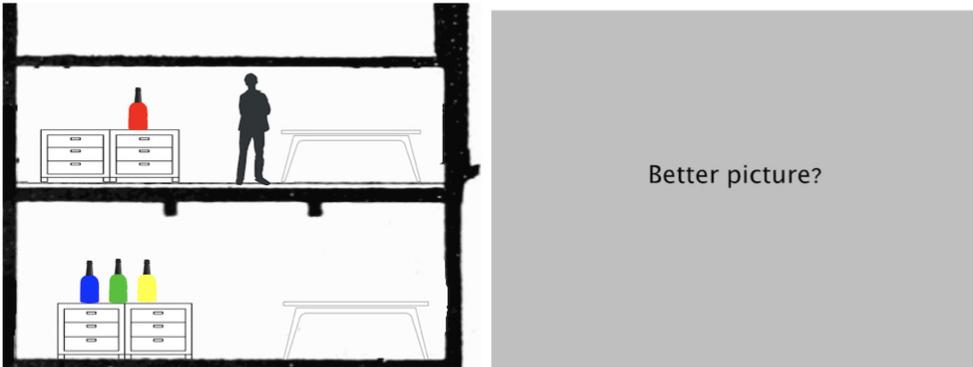


Figure 10: Exp 3 Number Filler Type I

**Number Filler Type II**

**There are six dogs**

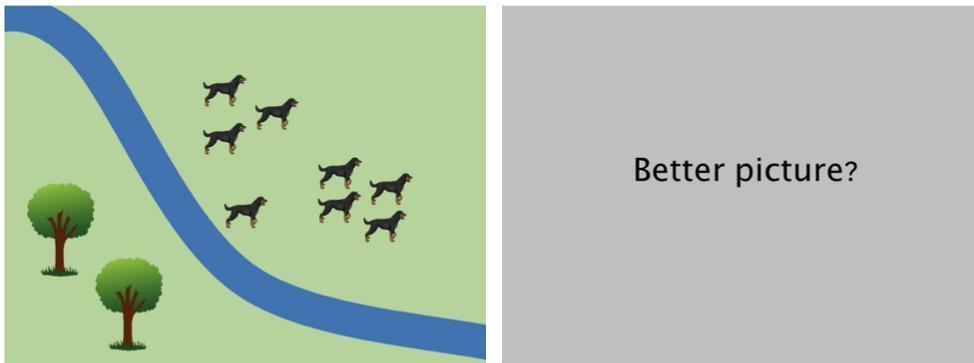


Figure 11: Exp 1 Number Filler Type II

There are three bottles on the upper floor

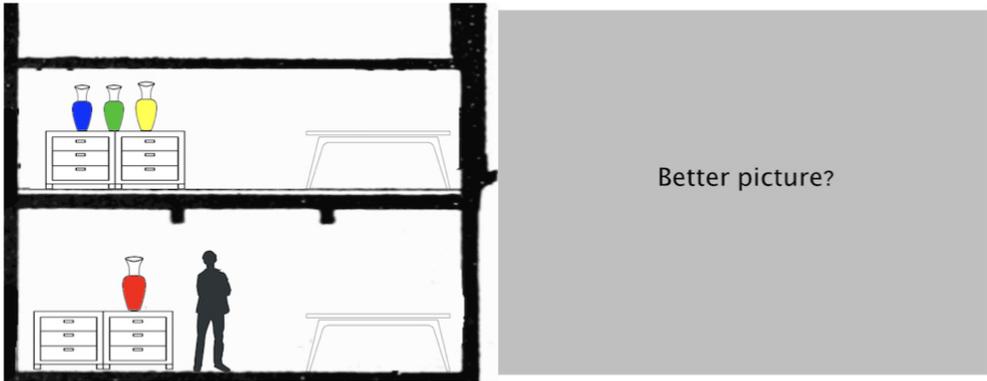


Figure 12: Exp 3 Number Filler Type II

**Number Filler Type III**

**There are six ponds**

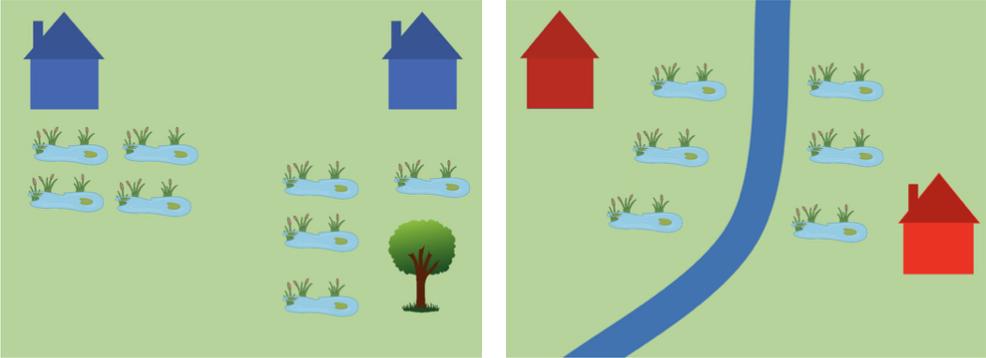


Figure 13: Exp 1 Number Filler Type III

There are three umbrellas on the lower floor

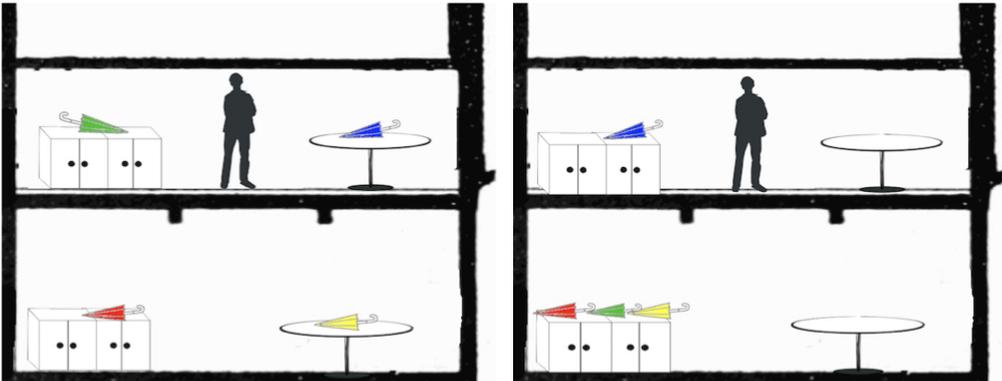


Figure 14: Exp 3 Number Filler Type III

7.2. Exp 1 Some Fillers

**Some Filler Type I**

**Some of the houses have a tree**

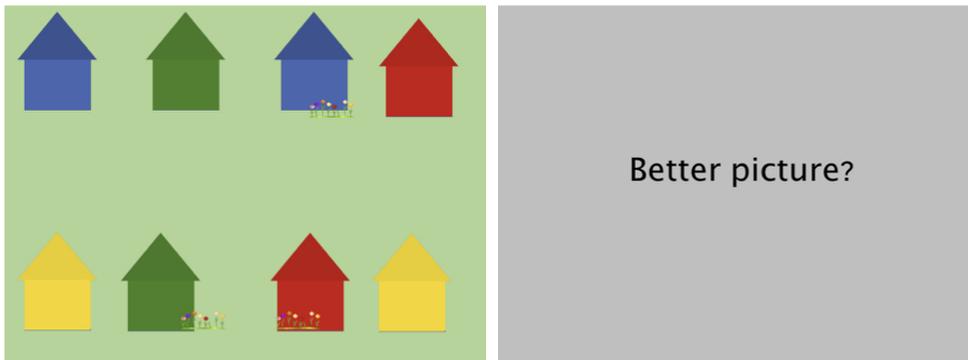


Figure 15: Exp 1 Some Filler Type I

All of the hammers are on the cabinets

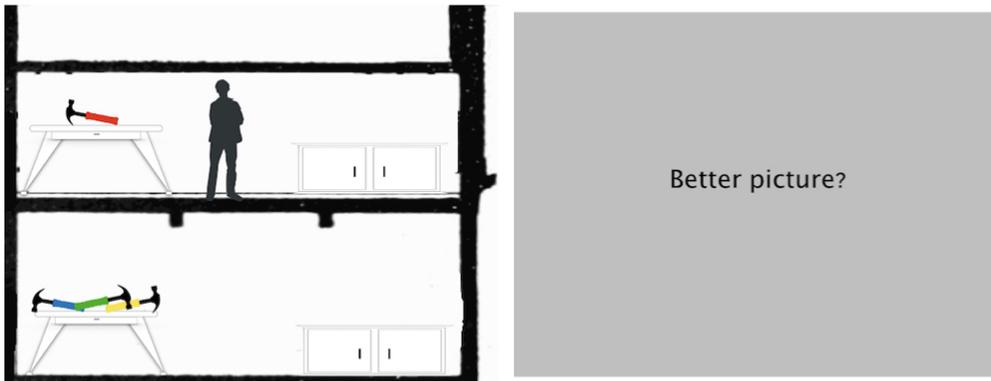


Figure 16: Exp 3 Some Filler Type I

**Some Filler Type II**

**All of the houses have a chimney**

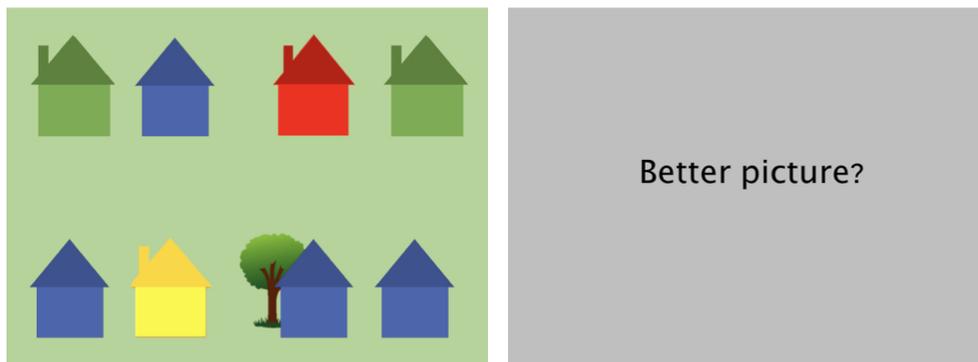


Figure 17: Exp 1 Some Filler Type II

All of the bottles are on the tables

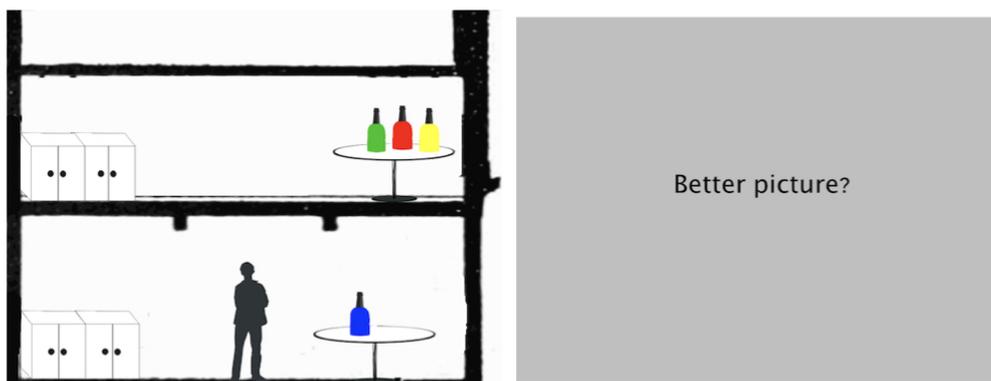


Figure 18: Exp 3 Some Filler Type II

**Some Filler Type III**

**All of the houses have a flower**

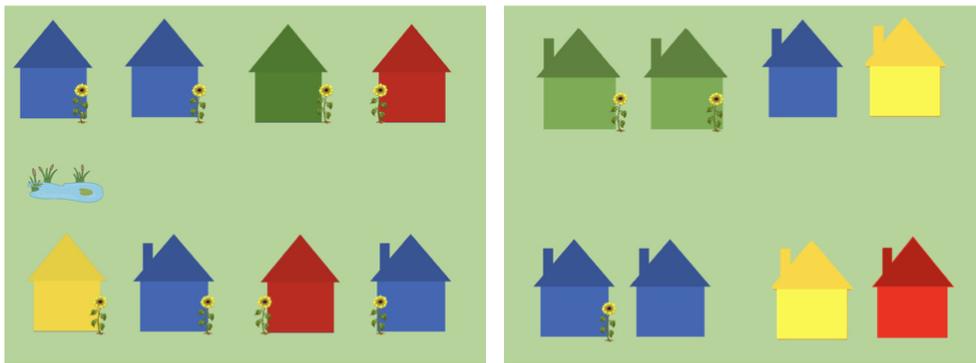


Figure 19: Exp 1 Some Filler Type III

All of the vases are on the cabinets

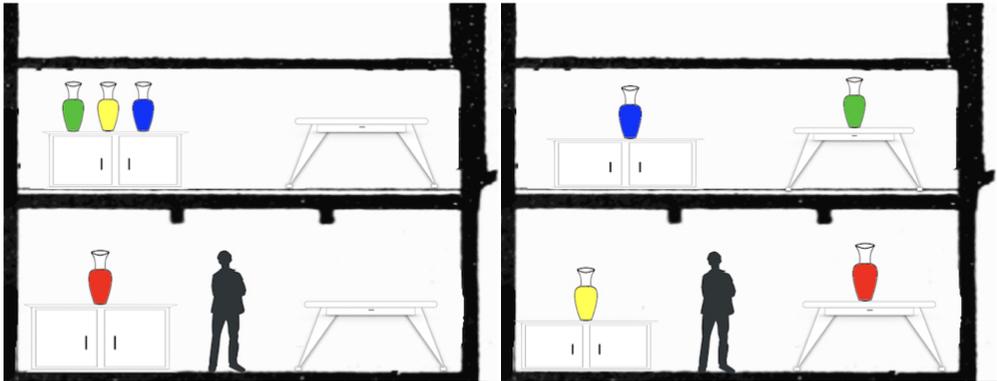


Figure 20: Exp 3 Some Filler Type III

7.3. Exp 1 Free Choice Fillers

**Free Choice Filler Type I**

**He can go to the yellow house but not to the red house**

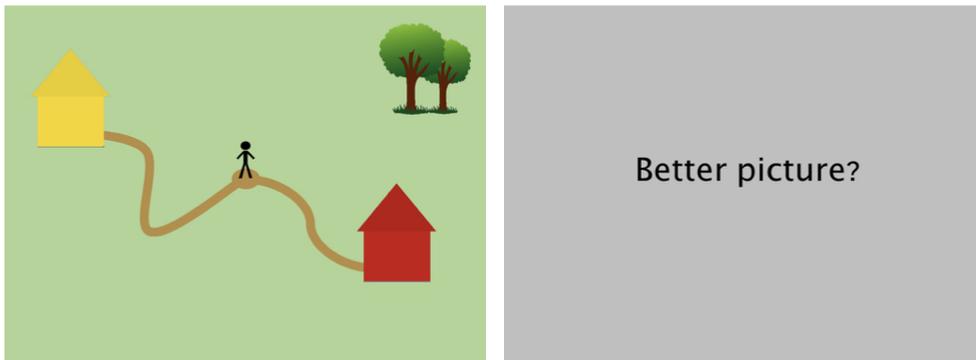


Figure 21: Exp 1 FC Filler Type I

He can pick up the red hammer but not the yellow hammer

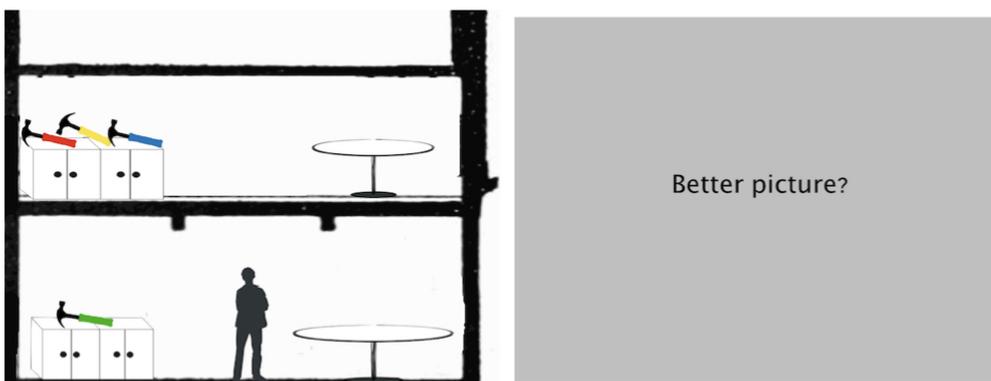


Figure 22: Exp 3 FC Filler Type I

**Free Choice Filler Type II**

**He can go to the yellow house but not to the red house**

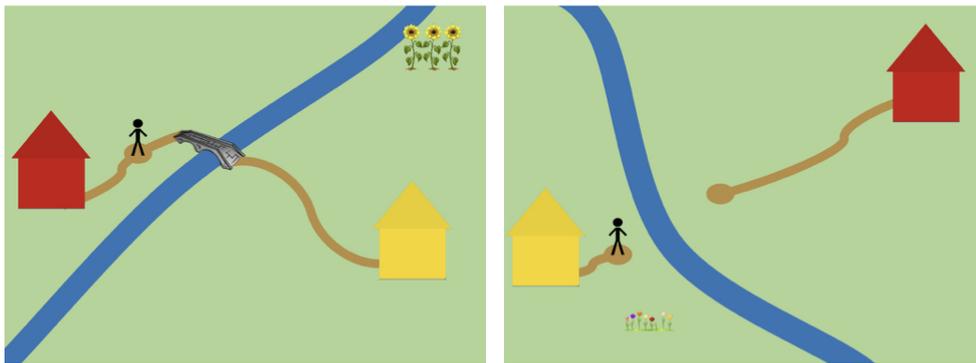


Figure 23: Exp 1 FC Filler Type II

He can pick up the blue bottle but not the green bottle

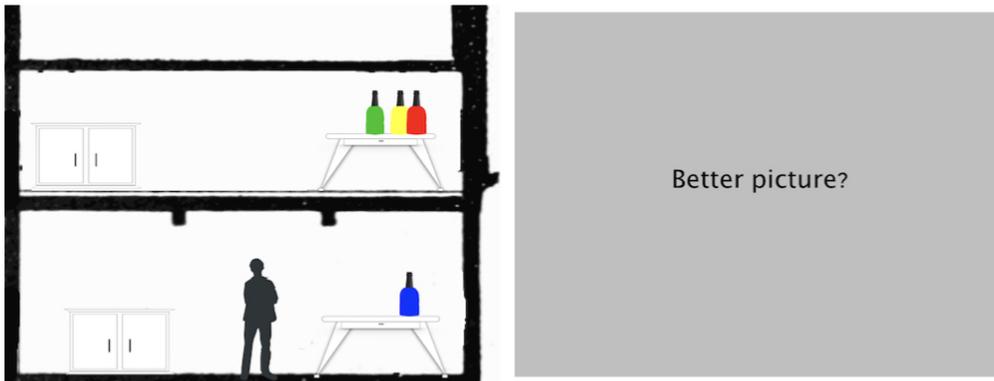


Figure 24: Exp 1 FC Filler Type II