

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository: <https://orca.cardiff.ac.uk/id/eprint/57925/>

This is the author's version of a work that was submitted to / accepted for publication.

Citation for final published version:

Chemla, Emmanuel and Bott, Lewis 2014. Processing inferences at the semantics/pragmatics frontier: Disjunctions and free choice. *Cognition* 130 (3) , pp. 380-396. 10.1016/j.cognition.2013.11.013

Publishers page: <http://dx.doi.org/10.1016/j.cognition.2013.11.013>

Please note:

Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher's version if you wish to cite this paper.

This version is being made available in accordance with publisher policies. See <http://orca.cf.ac.uk/policies.html> for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.



## Processing inferences at the semantics/pragmatics frontier: disjunctions and *free choice*

Emmanuel Chemla (LSCP, ENS, CNRS, France)

Lewis Bott (Cardiff University, UK)

[Please note that this is a postprint, but there be small differences between this and the version shown in Cognition]

**Abstract:** Linguistic inferences have traditionally been studied and categorized in several categories, such as entailments, implicatures or presuppositions. This typology is mostly based on traditional linguistic means, such as introspective judgments about phrases occurring in different constructions, in different conversational contexts. More recently, the processing properties of these inferences have also been studied (see, e.g., recent work showing that *scalar implicatures* is a costly phenomenon). Our focus is on *free choice permission*, a phenomenon by which conjunctive inferences are unexpectedly added to disjunctive sentences. For instance, a sentence such as “Mary is allowed to eat an ice-cream or a cake” is normally understood as granting permission both for eating an ice-cream *and* for eating a cake. We provide data from four processing studies, which show that, contrary to arguments coming from the theoretical literature, free choice inferences are different from scalar implicatures.

**Keywords:** pragmatics; processing; free choice; scalar implicatures; inferences

## 1. Introduction

The meaning we attach to any utterance comes from two sources. First, the specific combination of the words that were pronounced feeds the application of grammatical, compositional rules. The implicit knowledge of these rules allows us to understand any of the infinite combinations of words that form a proper sentence. Thus, grammatical operations lead to the literal meaning of the sentence. Second, the sentence meaning may be enriched by taking into account extra-linguistic information, such as general rules of communication and social interaction, information about the context of the utterance or the assumed common knowledge between speaker and addressee. The application of these pragmatic processes leads to the formation of implicatures. Grice (1975) illustrates this distinction with an example along the following lines. Consider a letter of recommendation for a teaching position that would read as in (1):

(1) Mr. Smith has beautiful handwriting, and he is neatly dressed at all times.

This letter puts forward positive features of Smith. Yet, when we go beyond the literal meaning of it, it is clear that this letter is destructive for Smith’s application. Letters of recommendation are supposed to deliver the most relevant and positive information about the applicant. If we take into account this pragmatic rule, it follows that Smith is a poor teacher. This inference is called an implicature of the utterance. In this example, the distinction is sharp: the literal meaning is entirely positive about Smith, but pragmatic principles add negative implicatures.

Implicatures have been traditionally studied by linguists and philosophers investigating

the divide between pragmatic principles and grammatical computations that allow people to derive the correct inferences (e.g., Ducrot, 1969; Grice, 1975; Horn, 1989; Levinson, 2000; Sperber & Wilson, 1995). More recently, psycholinguists have begun to investigate how various sorts of inferences are processed, adding a new kind of data to the debates, and with the ultimate goal of understanding how pragmatic principles might be instantiated in the human language processor. The most extensive investigations have been conducted on a kind of inference that we introduce in Section 1.1, so-called “scalar implicatures” (e.g., Bott & Noveck, 2004; Bott, Bailey & Grodner, 2012; Breheny, Katsos & Williams, 2006; Huang & Snedeker, 2008; Grodner, Klein et al, 2010; Nieuwland, Ditman, Kuperberg, 2010; Tomlinson, Bailey, & Bott, 2013), but work has also been completed on manner implicatures (Bott, Frisson & Murphy, 2008) and presuppositions (Chemla & Bott, 2013). The experiments in this paper investigate another type of implicature, free choice inferences, which arise from disjunctive sentences. An example is shown in (2) below.

(2) You can have chocolate cake or icecream.

Permissive disjunctions like this imply that both choices are possible, chocolate cake *and* icecream in this case. We will report several properties of these inferences that have been used to demonstrate that they do not belong to the core, literal meaning. Besides empirical properties, linguists have proposed an argument from parsimony: it has been shown that standard models for scalar implicatures can accommodate free choice inferences with little ado. Specifically, free choice inferences are often accounted for as second order scalar implicatures (see Section 1.3). This view leads to a very clear processing prediction that we will test: free choice should come at a processing cost similar or more extreme than any observed for scalar implicatures (see Section 1.1.2). The four current experiments test this prediction.

The significance of free choice inferences is twofold. First, from a theoretical perspective, it is natural to postulate a simple, boolean lexical entry for disjunctions. However, free choice challenges this assumption: conjunctive readings that arise in these contexts lack an explanation if natural language disjunction has a boolean disjunctive lexical meaning. We thus need to establish a convincing explanation for this phenomenon, and to motivate this explanation with independent empirical data. Second, free choice also plays a pivotal role in the typology of linguistic inferences. It has received both semantic and pragmatic accounts, and it has been assimilated to other widespread phenomena, like scalar implicatures (see section 1.3), of whose pragmatic or semantic status is at the core of the most lively debates in the theoretical literature. Hence, studies of free choice should inform us about semantics and pragmatics more broadly. Parallel arguments apply with respect to language processing. Understanding how free choice is processed is necessary if we are to understand processing of disjunction in general, and understanding whether we need separate processing mechanisms for *a priori* related phenomena informs us about the complexities of the linguistic processor.

We start by introducing some background about scalar implicatures and their relevant processing properties. In the following section, we present current arguments that lead to treat free choice as a pragmatic phenomenon. In this section, we will present the similarity between free choice and scalar implicatures, and derive from this comparison a processing prediction. Finally, we present three experiments that test whether free choice inferences and scalar implicatures show similar processing patterns.

## 1.1 Scalar implicatures

### 1.1.1 Derivation

Scalar implicatures are inferences that arise from the competition between an uttered sentence and a minimally different alternative that has not been uttered. Consider the example in (3). A sentence like (3) is naturally understood as implying that the sentence (4) is false. The reasoning supporting this inference relies on the assumption that, all things being equal, more informative sentences – like the alternative (4) – are to be preferred (Grice, 1975). The full derivation of the inference can be described in four steps.<sup>1</sup> (i) The literal meaning of (3) is computed. (ii) The listener notices that (4) is strictly more informative than (3): if John read all the books he surely read some, but not the other way round. (iii) Thus, all things being equal, it would have been preferable to utter (4) instead of (3). (iv) Something prevented the speaker from uttering (4), presumably that it is false; hence the implicature in (3-b).

(3) John read some of the books.

a. Literal meaning: One can find books that John read.

b. Implicature: John did not read all the books.

Enriched resulting meaning: John read some but not all the books.

(4) John read all the books.

### 1.1.2 The role of processing results

Much of the psycholinguistic research on implicatures has focused on whether scalar implicatures occur automatically and by default (chiefly motivated by arguments from Levinson, 2000). Although our own study is not directly concerned with this issue, the results that have been obtained in these studies motivate our own hypotheses and design. We will therefore briefly review some of the main findings (see also Katsos and Cummins, 2010).

According to default processing models of implicatures, scalar implicatures are derived on every occasion, and to interpret a sentence involving a scalar term without a scalar implicature requires cancelling the implicature component of the sentence. For example, to understand *some* to mean *at least one and possibly more*, requires first deriving the implicature associated with the scalar term, *not all*, and then cancelling this implicature, leaving only the literal meaning. This hypothesis was tested by Bott and Noveck (2004; see also Noveck & Posada, 2003) using a sentence verification task. Participants saw sentences that could be interpreted with or without the implicature. The sentences, e.g., *some elephants are mammals*, were such that whether the sentence is judged to be true depends on whether the participant derives the implicature. For example, with a scalar implicature the sentence is false, as in *some [but not all] elephants are mammals*. With the literal meaning on the other hand, the sentence is true, as in *some [and in fact all] elephants are mammals*. Bott and Noveck compared response times to the two types of responses and discovered that response times associated with implicatures were slower than to those with literal meanings, in direct contrast to the default model predictions. Bott, Bailey and Grodner (2012) expanded these experiments to demonstrate that the effects were not due to a speed-accuracy trade-off strategy and that they obtained across different languages, with different control sentences. Furthermore, Chevallier et al. (2008) showed that the same delay can be found with scalar

---

<sup>1</sup> See Spector (2003), van Rooij & Schulz (2004), Sauerland (2004), Francke (2011) and many others for refinements.

implicatures associated with other items (the exclusive reading of *or*, rather than the *not all* reading attached to *some*), which suggests that the processing pattern is quite generally associated with scalar implicatures. In short, these experiments demonstrate that there is a cost associated to verifying sentences with its implicature compared to verifying it without its implicature. This cost has been interpreted as a signature of the processes needed to derive a scalar implicature.

Similar findings have been revealed with other experimental techniques. For instance, Breheny, Katsos and Williams (2006) used self-paced reading experiments to show that the same phrase takes more time to process in a context in which it triggers a scalar implicature than in contexts in which it does not, and, Huang and Snedeker (2009) came to the same conclusion using a visual world paradigm (but see, Grodner, Klein, Carbury, & Tanenhaus, 2010). In short, there is a large range of evidence demonstrating that deriving scalar implicatures carries a cost. This cost has been used to argue against a default view of scalar implicature and, more generally, to answer questions about whether reasoning about alternatives should be understood as a late process or as an earlier pre-compiled operation. In the long run, these techniques may also be used to tease apart prominent approaches to scalar implicatures (e.g., Gricean, pragmatic approaches à la Spector, 2003, van Rooij & Schulz, 2004, Sauerland, 2004, Franke, 2011 and grammatical approaches à la Chierchia, 2004, Fox, 2007, Chierchia, Fox & Spector, 2008).

Our study was not designed to address such issues, however. Instead, we will target the comparison between scalar implicatures and free choice inferences. For these purposes we simply wish to establish that it takes time for scalar implicature interpretations to become available, and not be concerned with theoretical interpretations of these findings. In the following section, we will derive from this comparative approach a strong processing prediction for free choice inferences.

## 1.2 Free choice inferences

*Free choice* is the term used to describe a class of sentences in which disjunctions seem to be interpreted conjunctively (Kamp 1973, 1978). For instance, every sentence in (5)-(8) is disjunctive in nature (either because of the presence of a disjunction or because of the presence of a conjunction under a negation). Nonetheless, each of them gives rise to the *conjunction* of the inferences reported in the corresponding a and b items. In fact, these conjunctive inferences arise systematically when the disjunction is embedded in some specific linguistic environments. The case that will be most important is the one in (5) where a disjunction is embedded under an existential deontic modal (*is allowed*). This is also the source of the name of the phenomenon: by granting a disjunctive permission, it happens that free choice is given as to which of the two disjunctive alternatives is permissible.

- (5) Mary is allowed to eat an apple or a pear.
  - a.    ⇒ Mary is allowed to eat an apple.
  - b.    ⇒ Mary is allowed to eat a pear.
- (6) Bill may be in the kitchen or in the bathroom.
  - a.    ⇒ Bill may be in the kitchen.
  - b.    ⇒ Bill may be in the bathroom.
- (7) Mary does not have to do the homework and the reading.
  - a.    ⇒ Mary does not have to do the homework.
  - b.    ⇒ Mary does not have to do the reading.

(8) None of the students decided to take the in-class exam. Some students wrote an essay or did the homework.

- a.    ⇒ Some students wrote an essay.
- b.    ⇒ Some students did the homework.

The conjunctions of inferences in the examples above are not explained by the standard, Boolean interpretation of a disjunction. Consider the binary truth table associated with *or*. The sentence is true when one or other of the disjuncts is true; but it is not the case that both disjuncts<sup>2</sup> must be true to make the sentence true. Indeed, in most environments quite the opposite is implied. For example, *John speaks German or Dutch* normally implies that John does not speak both languages. Yet in the free choice examples above, the sentences are interpreted as allowing both “disjuncts” to be true. For example, *Mary is allowed to eat an apple or a pear* suggests that Mary is allowed to eat an apple *and* that Mary is allowed to eat a pear (whichever option she prefers).<sup>3</sup> The free choice interpretation therefore involves additional meaning beyond the standard disjunction.

Dominant accounts of free choice assimilate the phenomenon with scalar implicatures. These scalar implicature approaches are our main target, and we will discuss this class of theories in the next section, but we mention immediately that one of their advantages is that they allow us to maintain standard assumptions about disjunctions and modals. Some researchers have argued, however, that free choice is so robust that they warrant a reassessment of basic, semantic assumptions about disjunction and modals. Zimmermann (2000), Geurts (2005), Simons (2005), and Alonso-Ovalle (2008), for instance, base their analysis of free choice inferences on a re-analysis of disjunctions, as in *A or B*, not as boolean connectives, but as means to construct lists of objects ( $\{A, B\}$ ). Others start from a refinement of the contribution of the modal verb. For example, the accounts offered by Lewis (1979), Veltmann (1996) and van Rooij (2006) propose an explanation by trying to equate (semantically or pragmatically) *John is allowed to do A* with some conditional of the form *If John does A, then the law would be respected*.

Scalar implicature accounts of free choice thus rely on standard assumptions about the semantics of basic language units, while alternative accounts have to rely on modifications of these assumptions. We take this conservative aspect of scalar implicature accounts as an advantage. Let us make two additional general observations in favor of scalar implicature approaches. First, one can show that free choice inferences can be reproduced with expressions of disjunctions that do not involve actual disjunctions, as is the case when a conjunction occurs in the scope of negation (recall that *not(A and B)* is equivalent to *(not A) or (not B)*). Similarly, free choice inferences can be reproduced with a range of existential operators, beyond permissibility modals (see e.g., (8) above). Hence, if one had to account for free choice by modifying the semantics of disjunction or of modal verbs, the consequences on other instances of free choice would have to be investigated, which may involve revisiting the lexical semantics of other operators each time.

---

<sup>2</sup> We will use “disjunct” in a liberal manner: whenever a sentence contains a disjunction, we will refer to the sentence obtained by keeping one or the other of the two sides of the disjunction as a “disjunct”.

<sup>3</sup> We will not be concerned with the exclusive reading of the disjunction which may also arise in parallel of free choice, leading to the inference that if Mary is allowed to choose her preferred option (free choice), she cannot elect both options *at the same time* (exclusive reading). The two inferences are independent: one may arise without the other, and their source may be different. The second one plays no role in our experiments because it is systematically verified. We will stay away from this inference as much as possible to avoid confusion.

Second, free choice is cancellable, in a Gricean (and post-Gricean) sense. For example, although (5) is normally understood as implying (5)a and (5)b, this is not necessary and the following discourse is entirely natural: *Mary is allowed to eat an apple or a pear, although I can't remember which*. This type of cancellation can in principle be explained in many ways (see in particular Zimmermann, 2000, which discusses cancellation carefully as a signal of an ambiguity). However, it is difficult to disregard the similarity it creates between free choice and scalar implicatures (e.g., the status of the discourse above is similar to the following one in which the *not all* scalar implicature is cancelled: *Mary read some of the books, but I can't tell if she read them all*).

Overall, the arguments make a scalar implicature account of free choice both economical and plausible. Our goal in this study is to evaluate them on new grounds, namely the processing predictions they make. In the next section we explain how theories of scalar implicatures explain free choice.

### 1.3 Free choice inferences as scalar implicatures

Recently, linguists showed that formal models used to derive scalar implicatures can be adapted to account for free choice inferences (inspired by Kratzer and Shimoyama, 2002, see Schultz, 2005, Klinedinst, 2006, Fox, 2007, Chemla 2008, 2010, Alonso-Ovalle, 2008, Franke, 2011, for various implementations). In most cases, the idea is to treat free choice inferences as second order scalar implicatures in the following sense. Standard scalar implicatures enrich the meaning of an utterance by negating some alternatives. For example, if the speaker said *some*, then *all* can be negated, as in *John read some of the books*. Second-order scalar implicatures may enrich the meaning of an utterance by negating the enriched version of its alternatives. In other words, for what we call second-order implicatures, not only alternative sentences are considered, but the implicatures of these alternatives are also taken into account.

For example, consider the following derivation of free choice for (5) above, *Mary is allowed to eat an apple or a pear*. First, the alternatives of this sentence are given in (9). They are all stronger than (5), but we cannot negate both a and b consistently (there would be no permission left) and there would be no reason to negate one rather than the other. Hence, we would simply only negate the c alternative at this point: *Mary is not allowed to eat both an apple and an icecream at the same time* (see footnote 3).

- (9) a. Mary is allowed to eat an apple.
- b. Mary is allowed to eat a pear.
- c. Mary is allowed to eat an apple and a pear.

However, these alternatives may also be enriched with their own scalar implicatures. (9-a) with the negation of (9-b) becomes (10-a), and similarly (9-b) with the negation of (9-a) becomes (10-b). (For simplicity, we can ignore the last alternative (9-c) at this stage: it remains unchanged and plays no further role. This inference does play quite an important role, however, for unembedded disjunction. Schematically, the corresponding inference for “A or B” is “not (A and B)”, and it blocks the inferences “A” and “B” at a later stage).

- (10) a. Mary is allowed to eat an apple, but she's not allowed to eat a pear.
- b. Mary is allowed to eat a pear, but she's not allowed to eat an apple.

Now, we need to do some computations. (10-a) and (10-b) are stronger than (5), we thus get an overall meaning for the original sentence (5) with them negated as implicatures. We obtain the following: (i) Mary is allowed to eat an apple or a pear (literal meaning of (5)), but (ii) it's not the case that she's allowed to eat an apple and disallowed to eat a pear (negation of (10-a)), and (iii) it's not the case that she's allowed to eat a pear and disallowed to eat an apple (negation of (10-b)). This is equivalent to her being allowed to have any of the two (although possibly not the two of them at once, if we include the negation of (9-c)).

In a nutshell, the above theory derives free choice inferences as second-order scalar implicatures. It does not make any non-standard assumptions about disjunctions or other linguistic terms. It explains why free choice inferences are optional (just like other scalar implicatures). Finally, it can be shown to explain why free choice arises in some contexts, but not, for instance, with unembedded disjunctions.

## 1.4 Prediction

Let us take stock. We introduced scalar implicatures, an optional type of inference, and reported results showing that their derivation comes at some cost. We introduced free choice inferences, and explained that they could be accounted for as second order scalar implicatures. This view leads to a straightforward prediction about free choice inferences: if free choice inferences are “double” scalar implicatures, as suggested above, there should be an equal or greater cost to deriving free choice relative to the literal meaning (as there are with scalar implicatures). In short, free choice inferences should take longer to process than the simpler, logical disjunction. We next describe four experiments that test this hypothesis.

## 1.5 Overview of Experiments

Our experiments were based on the paradigm developed by Bott and Noveck (2004). Participants completed a sentence verification task in which they read sentences and responded with “true” or “false”. Crucially, the experimental sentences could be understood with either a free choice interpretation or a literal interpretation, just as sentences in Bott and Noveck could be understood with either a scalar implicature or a literal interpretation. The sentences were of the form, *X is allowed to save A or B*, where *X* was the name of a person and *A* and *B* were nouns. A cover story was constructed in which the destruction of the planet was described as imminent but that certain people were allowed to save certain types of objects. Specifically, zoologists were allowed to save living creatures and engineers were allowed to save artificial objects. An example of one of the sentences was, “Derek-the-engineer is allowed to save a hammer or a lion.” According to the cover story, a free choice interpretation would result in a false proposition (engineers are not allowed to save lions) whereas a literal interpretation would result in a true proposition (engineers are allowed to save hammers).

We conducted four experiments investigating the time course of free choice inferences. In each case we compared free choice *or* interpretations against literal *or* interpretations. Experiments 1, 2, and 4 were free choice versions of experiments from Bott and Noveck, (2004). In Experiment 1 participants were able to choose whether they responded with free choice or literal interpretations. In Experiment 2 we introduced a training context so that half the participants were biased towards a free choice response and half were biased towards a literal response. In Experiment 3, we added a direct test of scalar implicature to the free choice paradigm which allowed us to compare the two phenomena based on data from the same experimental run. Finally, in Experiment 4, we manipulated the time available for



participants to respond. If free choice inferences are an enriched form of the basic disjunction derived using similar procedures as scalar implicatures, strong free choice responses should be slower than weak, literal responses, just like strong scalar implicature responses are slower than corresponding weak, literal responses.

Our choice of methodology was guided by our goal: evaluate scalar implicature theory of free choice on processing grounds. Thus, we extended the seminal paradigm used to study the processing of scalar implicatures to study free choice. Indeed, while we have described the specifics of free choice and scalar implicature derivation, our main results are comparative and they will largely hold independently of what we assumed to be the underlying mechanisms behind either type of inference; all that matters for the present purposes is whether we observe the same pattern of results, that is, whether the processing mechanisms are similar or different.

## 2. Experiment 1

Participants classified experimental sentences of the form described above – true under their literal meaning, false under a free choice interpretation. We refer to the two readings as *literal* or *free choice* respectively.<sup>4</sup> In this experiment participants were not told which interpretation of the experimental sentences was “correct”; they simply chose the interpretation they felt was the most appropriate. We then compared response times for false response (corresponding to strong, free choice interpretations) against true responses (weak, literal interpretations). If free choice interpretations are derived in similar way as scalar implicatures, free choice interpretations should be delayed relative to the literal interpretations.

Participants also saw a range of control sentences to prevent strategies and ensure that the cover story was appropriately understood. The control sentences were either *double*, in which case they contained two nouns conjoined by *or*, or they were *single*, and contained only one noun. There were two phases to the experiment: a practice phase and an experimental phase. During the practice phase participants classified control sentences and received feedback indicating whether they were correct or incorrect. Experimental sentences were not presented in the practice phase. In the experimental phase, both types of sentences were presented but no feedback was provided.

### 2.1 Method

#### 2.1.1 Participants

Forty-six participants from Cardiff University completed the experiment for course credit or payment. Two were eliminated from the analysis because they failed to adequately respond to the control questions (see below).

#### 2.1.2 Design and materials

---

<sup>4</sup> There are two worries with this terminology that we would like to make explicit before using it. First, the word *literal* is biased towards the scalar implicature view. This view is our main target, however, and this terminology should thus help keep the relevant comparison between the two phenomena more visible. Second, even when one is concerned with scalar implicature, the description of the weak interpretation as the literal meaning is biased towards a Gricean account of the phenomenon, which is now highly debated. We use this terminology for simplicity and without committing ourselves to a contribution to this debate.

Participants saw sentences from three conditions, as illustrated in Table 1. First, they saw experimental sentences, in which the veracity varied depending on whether the participant derived the free choice or the weak interpretation. Experimental sentences always included two nouns separated by *or*. There were two experimental sentences per item; one in which the first disjunct was true and one in which the second disjunct was true, as in S1 and S2 in Table 1. (Here and elsewhere, we use a shortcut: “disjuncts” may refer to full sentences, obtained from a sentence containing a disjunction by replacing this disjunction to one or the other of its two sides. It is in that sense that a “disjunct” may be true or false). Second, double control sentences, which contained two nouns, just as in the experimental sentences, but the two disjuncts were either both true, as in S3, or both false, as in S4 (i.e., the nouns in the disjuncts were either both congruent or both incongruent with the special expertise introduced in the subject of the sentence, *zoologist* or *engineer*). Finally, there were single control sentences, which always contained only one noun. These could be either true, as in S5 or S6, or false, as in S7 or S8.

Condition	Sentence Number	Example	Correct response
Experimental	S1	Beverly-the-engineer is allowed to save a hammer or a lion.	T/F
	S2	Essie-the-engineer is allowed to save a kangaroo or a fork.	F/T
Double control	S3	Federico-the-engineer is allowed to save a hammer or a fork.	T
	S4	Martina-the-engineer is allowed to save a lion or a kangaroo.	F
Single control	S5	Maynard-the-engineer is allowed to save a hammer.	T
	S6	Alejandra-the-engineer is allowed to save a fork.	T
	S7	Cheryl-the-engineer is allowed to save a lion.	F
	S8	Rocco-the-engineer is allowed to save a kangaroo.	F

**Table 1.** Sentence types and examples.

Items consisted of quadruplets of four nouns, two of which were artifact nouns and two were living creatures. For example, the quadruplets hammer-fork-lion-kangaroo formed one item. To form the items, we compiled a list of 80 artifact nouns and 80 living creature nouns. From these, we selected 40 item bases consisting of two artifact nouns and two living creatures. Twenty of the quadruplets were arbitrarily assigned to form zoologist sentences and 20 to form engineer sentences. Each noun from a given quadruplets was therefore always paired with the name of an engineer/zoologist as appropriate throughout the experiment. Half of the item bases were allocated to the practice phase and half to the experimental phase.

Sentences were constructed by combining a name and item-appropriate profession as the subject, e.g., “Beverly-the-engineer” or “David-the-zoologist”, together with the verb segment, “is allowed to save a”, and one or two nouns. In the experimental phase, each noun from the item base was used to form one of the single control sentences, as in S5 to S8. The pair of artifact nouns and the pair of living creature nouns each formed one of the double control sentences, S3 and S4 respectively. Finally, one artifact noun and one living creature

noun was used to form one of the experimental sentences, S1, and the other pair of artifact/creature nouns formed the other experimental sentence, S2. In the practice phase, there were also 20 bases but only four sentences were formed from each base, the two double controls (S3 and S4) and two of the single controls (S5 and S7).

In the experimental phase then, there were 20 item bases and they formed 20 by 8 = 160 sentences. Of these, 40 were experimental sentences, 40 were double control sentences and 80 were single control sentences. In the practice phase there were 20 item bases but only 4 sentences from each base, hence only 80 sentences. All participants saw all items. The assignment of response keys to response options was randomized for each participant.

### *2.1.3 Procedure*

Participants were first presented with the cover story about the imminent destruction of the earth and its evacuation. They were told about engineers and zoologists, each of whom were allowed to take one and only one object away with them, corresponding to their specialty.<sup>5</sup> The crucial part of the cover story read, “A zoologist may save a living creature, whatever creature he or she wants. An engineer would not be allowed to save a living creature, but he or she may be allowed to save any kind of man-made, or artificial, object.” They were then told that they would see sentences on the screen and that they needed to say whether the sentences were true or false according to their general knowledge and the preceding scenario.

After reading the cover story participants progressed onto the practice phase of the experiment. Here, they read each sentence, made a response and then received corrective feedback. Sentences were presented in segments in the centre of the screen. There were seven segments for each sentence. The first segment was the name-profession combination, e.g., “Rocco-the-engineer”, which was presented for 750ms (250ms per word), the next five segments were the following five words in the sentence, each presented for 250ms, and the final segment was the noun or disjunction component, e.g., “lion” or “lion or kangaroo.” The final segment was left on screen until the participant made a response. RTs were taken from the onset of the final segment. Feedback consisted of the words, “correct” or “incorrect” as appropriate. After the practice phase participants were told to take a short break before commencing the experimental phase. The procedure was the same in the experimental phase except that feedback was not provided.

## **2.2 Results**

### *2.2.1 Data treatment*

Participants were removed if they failed to respond with 0.75 proportion correct on the control sentences. We reasoned that if they were not responding accurately to the control sentences they could have found alternative strategies to answer the experimental sentences. This criterion led to the removal of two participants.

We applied a log transformation to RTs to conform to the assumptions of parametric tests.

### *2.2.2 Inferential analysis*

---

<sup>5</sup> This particular part of the cover story was designed to make the independent “exclusive” inference driven by the disjunction true across all items, no matter what reading, literal or free choice, participants would draw (see footnote 3).

For all of the experiments presented in this article we combined items and participants as crossed random effects in a linear mixed-effects regression model (see Baayen, Davidson, & Bates, 2008) computed using the `lmer` function in R (Bates et al., 2011). Parametric models were assumed for RT data and logit models for choice proportions. We used models with random slopes and intercepts for participants/items when the factors were repeated measures, and random intercepts when they were between participants/items. This applied to main effects and interactions (see Barr, Levy, Scheepers, & Tily, 2013). On two occasions the maximal model failed to converge; we describe how we dealt with this in the text. We computed  $p$ -values by performing likelihood ratio tests in which the deviance of a model containing the fixed effect was compared with an otherwise identical model without the fixed effect.

### 2.2.3 Choice proportions

Control sentences were answered very accurately overall  $M = .93$ ,  $SD = .039$ , illustrating that participants understood the task and the cover story. In the experimental sentences, free choice responses (false) were the majority,  $M = .66$ ,  $SD = .33$ . The difference between the proportion of dominant (false) responses in the experimental sentences and the proportion accurate in the control sentences was significant,  $\chi^2(1) = 12.98$ ,  $p < .001$ ,  $\beta = 1.53$ ,  $SE = 0.37$ ,  $t = 4.14$ , suggesting that multiple interpretations were available for the experimental sentences.

We also examined the distribution of responses to the experimental sentences. Out of 44 participants, 4 responded entirely with free choice responses and 1 responded entirely with literal responses. The remaining participants exhibited a skewed distribution towards free choice responding (as suggested by the mean accuracy reported above).

### 2.2.4 Response Times

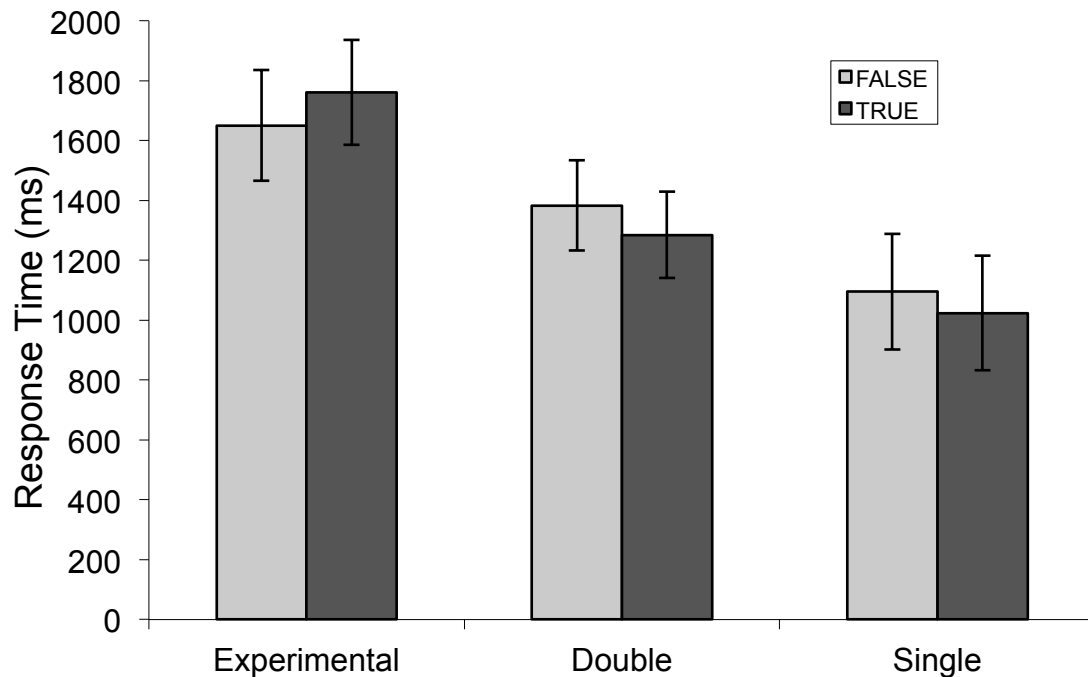
Figure 1 shows the pattern of RTs for the correct responses to all five types of sentences. For experimental sentences, false responses (free choice) appear faster than true responses but for the control sentences, the reverse is true.

An omnibus analysis comparing sentence type (single control, double control, experimental) against response type (true vs false)<sup>6</sup> revealed a significant interaction between sentence type and response type,  $\chi^2(2) = 14.62$ ,  $p < .001$ ,  $\beta = 0.054$ ,  $SE = 0.016$ ,  $t = 3.33$ . Analysis of the component effects revealed the significant interaction was driven by a smaller difference between true and false interpretations in the experimental sentences than between true and false in the control sentences. Specifically, there were no interactions between true and false responses and the double and single control sentences,  $\chi^2 < 1$ , but there were significant differences between response types when comparing the experimental sentences with the single control sentences,  $\chi^2(1) = 10.15$ ,  $p = .0014$ ,  $\beta = .054$ ,  $SE = 0.016$ ,  $t = 3.29$ , and the double control sentences,  $\chi^2(1) = 8.33$ ,  $p = .0039$ .

---

<sup>6</sup> The 3 by 2 maximal mixed model did not converge. We therefore removed the random effect that contributed least to the model fits and whose omission allowed convergence (following the recommendation in Barr et al., 2013). In this analysis it was the subject random slope for the sentence type by response type interaction for participants. The analysis from the restricted model is reported in the text. We also tested a model that combined the single and double control sentences into a single level and tested the 2 by 2 interaction of sentence type (experimental vs. control) by response type (true vs false). This model converged and the appropriate comparison was significant,  $\chi^2(1) = 11.98$ ,  $p < .001$ ,  $\beta = 0.058$ ,  $SE = 0.016$ ,  $t = 3.71$ .

In the experimental condition, the mean reaction time for free-choice response was numerically smaller than the mean for literal responses, but was not significant,  $\chi^2(1) = 1.23$ ,  $p = 0.27$ ,  $\beta = 0.037$ ,  $SE = 0.029$ ,  $t = 1.31$ . For control sentences true responses were faster than false responses,  $\chi^2(1) = 13.00$ ,  $p < .001$ ,  $\beta = 0.035$ ,  $SE = 0.0084$ ,  $t = 4.12$ .



**Figure 1.** Response times for Experiment 1. Strong, free choice interpretations are false responses to the experimental sentences; weak, literal interpretations are true responses to the experimental sentences. Control sentence results are RTs to correct responses. Error bars correspond to the standard error of the mean.

To make a more direct comparison between our results and those of Bott and Noveck (2004), we examined the confidence intervals around the difference between literal and free choice responses for our experimental sentences, and compared this against the reported difference in Bott and Noveck. We observed a 104ms advantage for free choice responses, with 95% (2-tailed) confidence intervals of [295ms, -87ms]. Thus, we can be 95% confident that free choice responses (the inference) were no more than 87ms slower than literal responses. In comparison, Bott and Noveck observed implicature responses (the inference) to be approximately 600ms slower than literal responses in Experiment 3 (the experiment with the most comparable design to this one).

## 2.3 Discussion

Experiment 1 was designed to test whether free choice interpretations were processed in a similar way to scalar implicatures. If so, we would have expected faster literal responses than free choice responses. We found no evidence that this was the case, however. On the contrary, relative to control sentences, we found significantly faster free choice responses than literal responses. Furthermore, confidence intervals indicated that the true difference between free choice and literal responses is unlikely to be of a comparable size and direction to those of scalar implicatures (Bott & Noveck, 2004).

Experiment 1 used a design in which participants were free to choose their interpretation of the experimental sentences. Allowing participants to choose, however, results in a design in which error responses are conflated with veridical responses. This introduces noise into the

analysis and consequently lowers the power of the experiment to detect differences across interpretations. Furthermore, the design is quasi-experimental, in that participants and items are not randomly allocated to levels of the experimental factor. In Experiment 2 we used a cleaner design in which participants were presented with a context that could be either biased towards a free choice response or a literal response (similar to Experiment 1 of Bott & Noveck, 2004).

### 3. Experiment 2

In Experiment 2 participants also classified sentences involving zoologists and engineers. However, participants now received feedback on the experimental sentences as well as the control sentences. One group of participants received corrective feedback encouraging the free choice interpretation of the experimental sentences, and the other group received feedback encouraging the literal interpretation. Thus, participants who were in the free choice group were trained to respond “false” to the experimental sentences whereas those in the literal group were trained to respond “true”. The feedback occurred in the practice phase at the same time as participants received feedback on the control sentences. The comparison of interest was between true responses to the experimental sentences (from the literal group) and false responses to the experimental sentences (from the free choice group).<sup>7</sup>

#### 3.1 Method

##### 3.1.1 Participants

Sixty-seven participants from Cardiff University completed the experiment for course credit or payment. Seven participants were removed from the analysis because they failed to reach the inclusion criterion of 0.75 proportion correct on the control sentences. Four of these were from the literal group and three from the free choice group.

##### 3.1.2 Design, materials and procedure.

---

<sup>7</sup> If training was completely effective in determining an interpretation for a particular kind of structure, participants from the literal group and from the free choice group should have a literal interpretation and a free choice interpretation, respectively, not only for the experimental sentences but for the control sentences as well. The only difference between control and experimental sentences then is that true/false responses for the controls do not discriminate between the two possible interpretations. As a result, any RT effect of interpretation should be visible between groups both in experimental sentences and in control sentences. What we find, however, is that response patterns for control and experimental sentences are different. We believe that this protects our interpretation of the results and justify *a posteriori* our use of control sentences. We mention two possible interpretations for the absence of interpretation effects on control and experimental results.

(i) In experimental conditions we were able to clean the data corresponding to 'wrong' interpretations (e.g., a free choice interpretation by a participant in the literal group would lead to a detectable false response and can be excluded). Hence, if anything, an effect of interpretation will be less polluted by noise in the analysis of experimental sentences. In the control conditions, however, there might be some proportions of 'wrong' interpretations, which would still lead to correct responses. As a result, the data from the control sentences can serve as a baseline for our effect, by virtue of being an impure version of our experimental conditions. Note that the same argument applies to experiments without training: there is even less ground to decide which interpretation participants adopted for the control conditions, so that these conditions can be used as the same type of impure baseline as in the experiments with training.

(ii) Our results may be interpreted as evidence for the fact that there might be a coarse grained semantic analysis in this task so that, in control conditions, participants are able to detect that both interpretations lead to the same response without deriving any of these interpretations. Hence, they would be able to provide an answer without going through the derivation of any interpretation, at least in some occasions, and any effect of interpretation should be obscured. This interpretation also applies to experiments without training.

Participants underwent a training phase and an experimental phase, just as they did in Experiment 1. The difference was that during the training phase of Experiment 2, participants received feedback on the experimental as well as the control sentences. Participants from the free choice group received feedback indicating that the correct response to the experimental sentences was “false”, whereas participants from the literal group received feedback indicating that it was “true”. Feedback was corrective (“incorrect” or “correct”) but did not include a rationale. For example, if a participant from the free choice group responded “true” to an experimental sentence, the word, “incorrect” appeared on the screen. There was no difference between the feedback provided about responses to experimental sentences than to feedback about control sentences.

Participants saw the same number of control trials as in Experiment 1 (80) but they also saw 20 experimental sentences. The experimental phase and all other aspects of the design were identical to that of Experiment 1.

## 3.2 Results

### 3.2.1 Choice Proportions

Accuracy was high and approximately equal across conditions, both for the control sentences,  $M_{fc} = .93$ ,  $SD = .043$ , vs  $M_{lit} = .92$ ,  $SD = .036$ , and for the experimental sentences,  $M_{fc} = .95$ ,  $SD = .039$ , vs  $M_{lit} = .95$ ,  $SD = .049$ . The training used to bias responses was therefore effective in both groups.

### 3.2.2 Response times

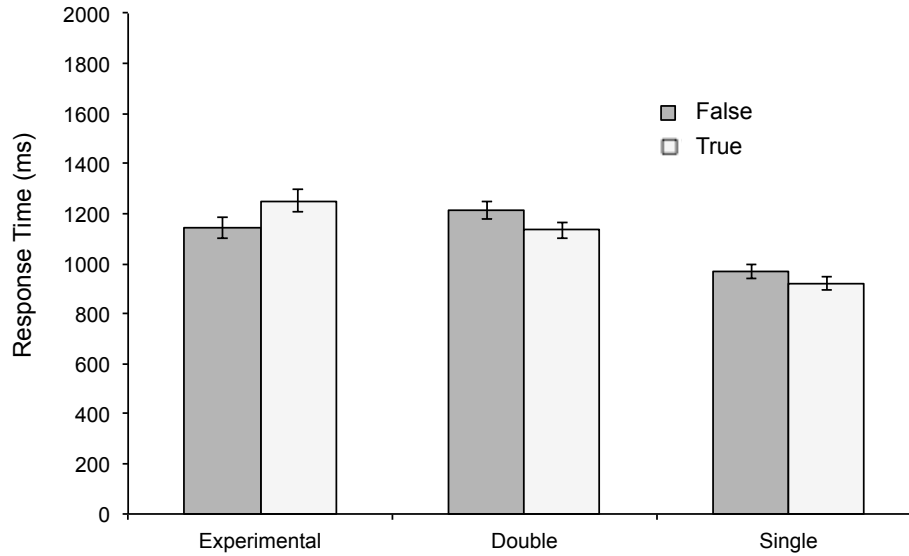
Figure 2 shows the pattern of RTs. For the experimental sentences, false (free choice) responses appear faster than true (literal) responses, but the reverse applies for the control sentences. The pattern therefore replicates Experiment 1.

The omnibus analysis revealed a significant interaction between sentence type (control vs. experimental) and response (true or false)<sup>8</sup>,  $\chi^2(2) = 25.00$ ,  $p < .001$ ,  $\beta = 0.012$ ,  $SE = 0.0081$ ,  $t = 4.97$ . For control sentences, true responses were significantly faster than false responses,  $\chi^2(1) = 8.90$ ,  $p = .0029$ ,  $\beta = 0.028$ ,  $SE = .0089$ ,  $t = 3.16$ , just as in Experiment 1, but for experimental sentences, false responses (free choice) were derived significantly faster than true (literal) responses,  $\chi^2(1) = 23.89$ ,  $p < .001$ ,  $\beta = 0.12$ ,  $SE = .024$ ,  $t = 4.94$ .

---

<sup>8</sup> The 3 by 2 maximal mixed model failed to converge. We followed the procedure we used in Experiment 1 and removed the random effect that contributed least to the model fit and whose removal led to convergence. In this case the item random slope for the sentence type by response interaction was removed. We also verified that the maximal model when the control conditions were combined into a single level was also significant,  $\chi^2(1) = 11.42$ ,  $p < .001$ ,  $\beta = 0.047$ ,  $SE = 0.0096$ ,  $t = 4.90$ , just as we did in Experiment 1.





**Figure 2.** Response times for Experiment 2. Free choice and literal results (respectively false and true responses to the experimental sentences) are the correct responses from the free choice and literal participants respectively. Control sentence results (double and single) are RTs to correct responses averaged across both groups of participants. Error bars correspond to the standard error of the mean.

### 3.3 Discussion

In Experiment 2 we biased participants towards interpreting the experimental sentences as either literal or free choice. The design had greater power than Experiment 1 and involved a complete experimental design in which items and participants were randomly allocated to experimental levels. We observed significantly faster free choice responses than literal responses, in contrast to a scalar implicature account of free choice.

One limitation to our conclusions so far is that they have been based entirely on cross-experimental comparisons: we have assumed that the results from our experiments are comparable to results from other sentence verification paradigms that tested scalar implicatures. This assumption may not be justified, however. Note for instance that our free choice verification task was based on a cover story, while standard implicature verification tasks are based on world knowledge. This difference allows for the possibility that scalar implicatures computed under our experimental conditions may not exhibit delayed responses compared to literal interpretations. In Experiment 3 we sought to remove this limitation by directly comparing scalar implicatures and free choice interpretations within the same experiment.

## 4. Experiment 3

In Experiment 3 participants undertook two sentence verification tasks in two experimental blocks. In one block, participants completed a scalar implicature task, whereas in the other block they completed a free choice task (with counter-balanced order). Our primary aim in this experiment was to test the interaction between weak and strong interpretations across the two phenomena, free choice and scalar implicature. We refer to the weak meaning as the *literal* meaning, and to the strong meaning (free choice or implicature, respectively) as the *inference* meaning.

The free choice task was exactly the same as Experiment 2. The scalar task was similar to Bott and Noveck (2004) but modified to make it more comparable to the free choice task. In particular, the task now involved a cover story against which the truth values of the target sentences should be evaluated. As a result, we also avoided using the supercategory-subcategory sentence frame (e.g., “some elephants are mammals”), which arguably creates odd sentences (see Magri, 2009, for theoretical discussions of conditions under which scalar implicatures in contradiction with common knowledge may create oddness).

We chose to test the usual *not all* inference of *some*. There has been a suggestion that a more direct test would have been to test the exclusive inference associated with the disjunction (*or*). However, simple disjunctive sentences come with additional inferences (ignorance inferences about the disjunct, often called primary scalar implicatures, e.g., Spector 2003, Sauerland 2004, van Rooij and Schulz 2004) and some theorists have even disputed the possibility that this inference is a scalar implicature (e.g., Geurts 2005). Hence, the exclusive reading of simple disjunctions does not qualify as an appropriate, uncontroversial baseline for scalar implicature processing. The fact that it involves the same word (*or*) makes it a better candidate for superficial reasons only.

Based on Bott and Noveck (2004)’s seminal idea, target sentences were true under their literal reading, and false under their inference reading. Whether a participant endorsed the literal or the implicature interpretation was determined by corrective feedback, in a similar way as in Experiment 2. One group of participants, the *literal* group, made true responses to the experimental sentences, and other participants, an *inference* group, made false responses to the experimental sentences. Participants were consistently literal (or inference) in both tasks, that is, the same set of participants made literal (or inference) responses in the free choice and the scalar tasks. If the results from the scalar task are consistent with previous studies, responses from the inference group should be longer than those from the literal group. More originally, if free choice inferences are double scalar implicatures, we should find a larger difference for free choice sentences, contrary to what we found in Experiments 1 and 2.

## 4.1 Participants

Forty-four Cardiff University students participated for course credit. Participants data were removed from a given task if accuracy on the control conditions was below 0.75 proportion correct (as in Experiments 1 and 2). The data from five participants were removed from the scalar task (two from the literal group) and seven from the free choice task (five from the literal group).

## 4.2 Design and Materials.

The design for the FC task was identical to Experiment 2. The design for the scalar task was as follows.

Participants made truth value judgments about sentences that involved a quantifier (*some* or *all*), an exemplar (either a bird or a plain terrestrial animal, henceforth an animal), and a verb (*saved*, *killed*, or *infected*). All sentences were of the form “[Quantifier] [exemplar] were [verb].” There was one type of experimental sentence and eight types of control sentence. Examples of each are shown in Table 2. Each of the 20 experimental sentences involved a different bird, and the birds assigned to the experimental sentences were not used in the control conditions. The birds and animals in the control sentences were drawn from a bank of 35 items with some repetition of items across conditions.

Label	Example	Truth	Frequency
Some-bird-saved [EXP]	Some eagles were saved.	T/F	20
some-bird-killed	Some starlings were killed.	F	15
some-bird-infected	Some blackbirds were infected.	T	15
all-bird-saved	All ducks were saved.	T	10
all-bird-killed	All hens were killed.	F	10
all-bird-infected	All robins were infected.	F	10
some-animal-killed	Some antelopes were killed.	T	5
some-animal-infected	Some foxes were infected.	T	5
all-animal-saved	All ponies were saved.	F	10

**Table 2.** Sentence types and examples.

The truth value of the sentences was determined by a cover story. This described a scenario in which a new virus had infected birds and animals. The important parts of the cover story were that (i) the virus infected only males (females were resistant), and (ii) the virus killed only animals (birds always survived). Consideration of the cover story reveals how the truth values shown in Table 2 are derived. For example, *some starlings were infected* is true because male starlings caught the virus. Crucially, whether the sentence is judged to be true depends on whether the scalar implicature was derived. For example, *some eagles were saved* could be read as true (because at least one eagle survived), or false (because all eagles survived).

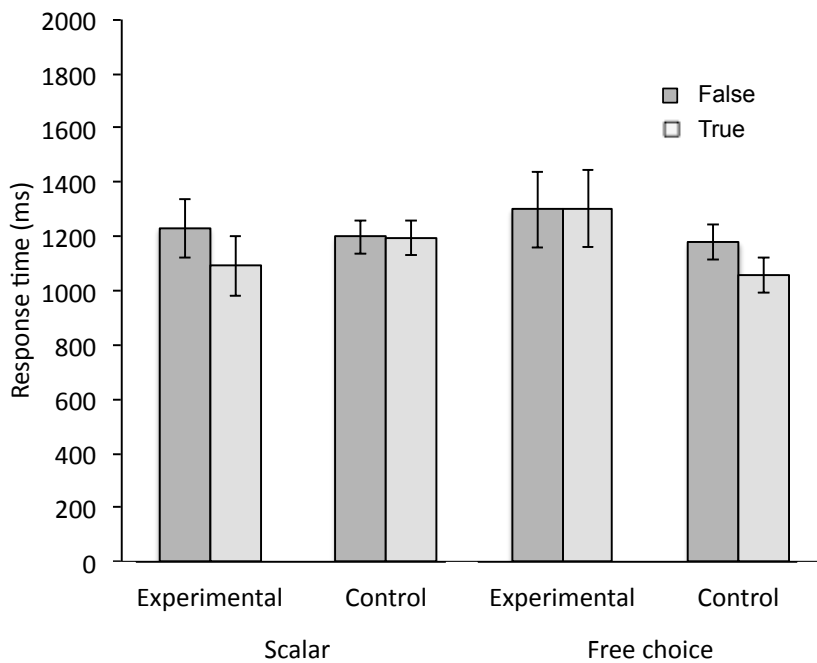
The control sentences were chosen to prevent participants predicting the correct answer to the experimental sentence without reading or combining the different segments. Thus, it was not possible to accurately predict whether the experimental sentences were true or false by relying on only the first part of the sentence, “Some [bird] ...,” or only on the last part of the sentence, “... were saved.” The control sentences were of eight types. These were formed by varying the quantifier (*some* vs *all*), whether the subject was a bird or an animal, and whether the verb was *saved*, *killed*, or *infected*. Overall there were more false control sentences than true control sentences (45:35) (this might have introduced a small bias towards false responding in this task but, to forecast our results somewhat, we found facilitation of true response, not false responses, and hence the impact of this imbalance was minimal).

The task included a practice phase and an experimental phase. During the practice phase, participants received feedback on their responses, but during the experimental phase they did not (just as in Experiments 1 and 2). The sentences were of the same form as those in the experimental phase but involved different items. Participants judged 40 sentences (half the number of the sentences in the experimental phase), broken down in approximately the same ratios as the experimental phase, e.g., there were ten experimental sentences, seven *some-bird-killed* sentences, and eight *some-bird-infected* sentences. The interpretation of the experimental sentence was determined by the feedback provided in the practice phase. For

*literal* participants, feedback indicated that the experimental sentences were true, whereas for *inference* participants, feedback indicated that the sentences were false.

Participants completed the free choice task and the scalar task. Participants were randomly assigned to a literal condition, in which case they completed the literal conditions of the free choice task and the scalar task, or the inference condition, in which case they completed the inference conditions of both tasks. The order in which participants completed the tasks was counter-balanced.

Participants were told that they would be doing two sentence verification tasks but not that the two tasks were linked. After the first task was completed the experimenter started the next test by executing a different computer program. It was only after both tasks had ended, during the debriefing stage, that participants were told that there was a link between the two.



**Figure 3.** Raw response times for the scalar task and the free choice task, Experiment 3. Incorrect responses have been removed. In the experimental conditions, true responses correspond to literal interpretations and false responses to inference interpretations. Error bars are the standard error of the mean.

### 4.3 Scalar task results

#### 4.3.1 Choice proportions

Performance was accurate across both training groups and differed very little. For the control sentences,  $M_{lit} = .93$ ,  $SD = .036$ , vs  $M_{inf} = .95$ ,  $SD = .076$ , and for the experimental sentences,  $M_{lit} = .95$ ,  $SD = .076$ , vs  $M_{inf} = .94$ ,  $SD = .096$ . The training used to bias responses was therefore effective in both groups.

#### 4.3.2 Response times

All incorrect responses and outliers were removed from the data analysis. Responses were considered outliers if they were greater than 10s. There were 12 outliers out of 3900 responses.

The left half of Figure 3 shows the results from the scalar implicature task. Participants appear to need longer to respond to the scalar inference (false) compared to the literal meaning, consistent with Bott and Noveck (2004) and others, but there is very little difference between the control conditions. This pattern was confirmed by a significant response type (true vs false) by sentence type (control vs experimental) interaction,  $\chi^2(1) = 9.30, p = .0023, \beta = 0.068, SE = 0.022, t = 3.17$ .

There was no significant difference between true (literal) vs false (scalar inference) on the experimental sentences,  $\chi^2(1) = 1.32, p = .25, \beta = 0.041, SE = .035, t = 1.19$ , but this might be expected because training was a between participants manipulation and we were consequently unable to subtract away individual participant variability (Bott & Noveck, 2004, used a within-subject design, and therefore did not suffer the same problem). The interaction we report above, in contrast, removes individual participant variability by estimating it from the control conditions. No significant difference was observed between true and false control conditions,  $\chi^2(1) = 1.12, p = .29, \beta = 0.011, SE = .011, t = 1.06^9$ .

We also examined whether the order in which the participants completed the scalar task and free choice task had an effect on the results. No effects of order were apparent, however,  $\chi^2$ 's < 1.

### 4.4 Free choice task results

#### 4.4.1 Choice proportions

Accuracy was good overall and similar to Experiment 2 for the control sentences ( $M_{\text{inf}} = .93, SD = .045$ , vs  $M_{\text{lit}} = .95, SD = .039$ ) and for the experimental sentences ( $M_{\text{inf}} = .93, SD = .079$ , vs  $M_{\text{lit}} = .97, SD = .035$ ).

#### 4.4.2 Response times

Incorrect responses and outliers were removed. There were two outliers.

As with the scalar task, this resulted in a significant 2 by 2 interaction of response type (true vs false) and sentence type (control vs experimental sentences),  $\chi^2(1) = 6.05, p = .014, \beta = 0.040, SE = 0.013, t = 3.06$ . Analysis of the simple effects revealed no significant difference between false (free choice) and true (literal) for the experimental sentences,  $\chi^2 < 1$ , but there was a significant difference between true and false control conditions,  $\chi^2(1) = 19.49, p < .001, \beta = 0.056, SE = 0.010, t = 5.48$ .

No main effects or interactions with order were present,  $\chi^2$ 's < 1.

---

<sup>9</sup> This could be considered unusual given the well-known bias against responding false. The most likely explanation for our failure to observe this effect is that we had a predominance of false control sentences in the scalar task (see the Method section), which offset any general bias against false responding. Note that any bias towards false responding would reduce the observed cost of a scalar implicature response, and so would work in the opposite direction to our hypothesis.

## 4.5. Comparison between scalar and free choice

The main aim of the experiment was to compare the literal and inference responses across scalar and free choice tasks. The significant interactions within the scalar task and the free choice task suggests a three-way interaction between task (scalar vs free choice), sentence type (control vs experimental sentence) and interpretation (literal vs free choice). This was indeed the case:  $\chi^2(1) = 7.028, p < .005, \beta = 0.086, SE = 0.028, t = 3.08$ . The 2 by 2 interaction between task (scalar vs free choice) and interpretation (literal vs free choice) restricted to experimental conditions was not significant,  $\chi^2 < 1$ , however.

To gain some statistical power, we conducted cross-experimental analyses between the scalar task of Experiment 3 and the free choice tasks of Experiments 2. While participants were clearly not assigned randomly across experiments, there is little chance of sampling bias determining our results because (i) we were testing interactions and not main effects, and (ii) participants from both studies were from similar samples (i.e. students from Cardiff University). This analysis revealed a significant three-way interaction between task (scalar vs free choice), sentence type (control vs experimental sentence) and interpretation (literal vs free choice),  $\chi^2(1) = 23.56, p < .001, \beta = 0.075, SE = 0.018, t = 4.15$ , just as we found in Experiment 3. Furthermore, the 2 by 2 task (scalar vs free choice) by interpretation (literal vs free choice) interaction, restricted to experimental conditions, was then also significant,  $\chi^2(1) = 9.65, p = .0033, \beta = 0.044, SE = 0.015, t = 2.97$

There were no effects of order involving interpretation in Experiment 3,  $\chi^2(1) < 2.40, p's > .31$ , nor a main effect of order,  $\chi^2 < 1$ . However, there was an order by task interaction,  $\chi^2(1) = 6.38, p = .011, \beta = 0.052, SE = 0.020, t = 2.71$ , such that participants completed the scalar task more quickly when they completed it second compared to when they completed it first, but no effect was observed for the free choice task.

## 4.6 Discussion

In Experiment 3 we found that scalar implicature interpretations needed more time to verify than corresponding literal interpretations, relative to the control sentences. This result replicates previous findings but with a new design using a cover story and more natural sounding sentences. We also found that free choice interpretations were verified just as quickly as corresponding literal interpretations. Most importantly, we observed a significant interaction across task, interpretation and control conditions, demonstrating that the difference between the inference and literal interpretations for scalar implicature sentences was greater than the difference between inference and literal interpretations for free choice sentences, relative to control sentences. Significant interactions were also observed between the scalar task of Experiment 3 and the free choice task of Experiment 2. These results are inconsistent with a processing model of free choice that assumes that free choice inferences are particularly complex, or double, scalar implicatures. Such an account would predict a greater cost to deriving free choice inferences than scalar implicatures.

## 5. Experiment 4

In our final experiment we tested whether the proportion of free choice inferences was reduced when participants had a small time window in which to process the sentences. Our design follows Bott and Noveck (2004), Experiment 4, in which participants received either 900ms (short lag) or 3000ms (long lag) to verify the sentences. Bott and Noveck observed that the proportion of scalar implicature responses increased from 0.28 in the short lag to 0.44

in the long lag. If free choice inferences are complex scalar implicatures, the reduction in free choice inferences from the long lag to the short lag should be at least as much as that observed by Bott and Noveck.

Participants performed a sentence verification task similar to Experiment 1. The crucial difference was that participants were required to respond at a specific time after the onset of the final sentence segment (e.g., “a hammer or a lion”). Participants saw the sentence presented in the centre of the screen and then waited until a set of exclamation marks appeared (the deadline, “!!!!”). They then had to make their response immediately after the deadline. For one group of participants the deadline occurred after 1000ms (the short lag group), and for the other group of participants, the deadline occurred after 3000ms (the long lag group).

The variable of interest in this experiment was the change in proportion of inference responses across the lag manipulation, just as it was in Bott and Noveck (2004). Participants were therefore not given feedback on experimental sentences in the practice phase (c.f. Experiments 2 and 3).

## 5.1 Method

### 5.1.1 Participants

Eighty-six participants completed the study for course credit or payment. Seven participants were removed for having poor accuracy or timing performance (see below). Forty participants remained in the short-lag condition and 39 in the long lag condition.

### 5.1.2 Design, Materials and Procedure

There were three phases to the experiment. The first was a practice phase, similar to that of Experiment 1, in which participants judged control sentences and received feedback on whether they were correct. This allowed them to consolidate the cover story and get general practice with the task. The second was a practice phase to familiarize participants with the deadline procedure, and the third was the experimental phase.

Participants saw sentences presented in the centre of the screen in the same manner as Experiment 1. During the deadline practice phase and the experimental phase, participants were told to respond within 500ms of the signal and not before the signal. On each trial they received one of three types of feedback. If their response was on time, the response time was displayed on the screen, e.g., “235ms”. If they had made an anticipatory response (before the signal or up to 100ms after the signal), they saw a message reading, “too quick”, and if they had responded late (more than 500ms after the signal), they saw a message reading, “too slow.” They did not receive feedback on whether the true/false response was correct or incorrect.

Participants judged double control sentences (true and false) and experimental sentences (see Table 1). We did not include single control sentences because they contained fewer words than the other types of sentences and the effect of the lag manipulation would have been difficult to interpret. Forty double control sentences were used in the first practice phase and 40 in the deadline practice phase. There were 80 sentences in the experimental phase, 40 of which were double control sentences and 40 were experimental sentences. The sentences were constructed in the same way as in Experiment 1.

## 5.2 Results

### 5.2.1 Preprocessing

Participants were removed from the analysis if they responded with less than 50% responses in time (3 participants) or if they were less than 50% accurate in the control conditions (5 participants).

We also removed all responses that were outside the response window (100ms-500ms after the deadline signal). This accounted for 17% of the data. The qualitative conclusions of the experiment are similar regardless of whether this data is excluded, however.

### 5.2.2 Analysis

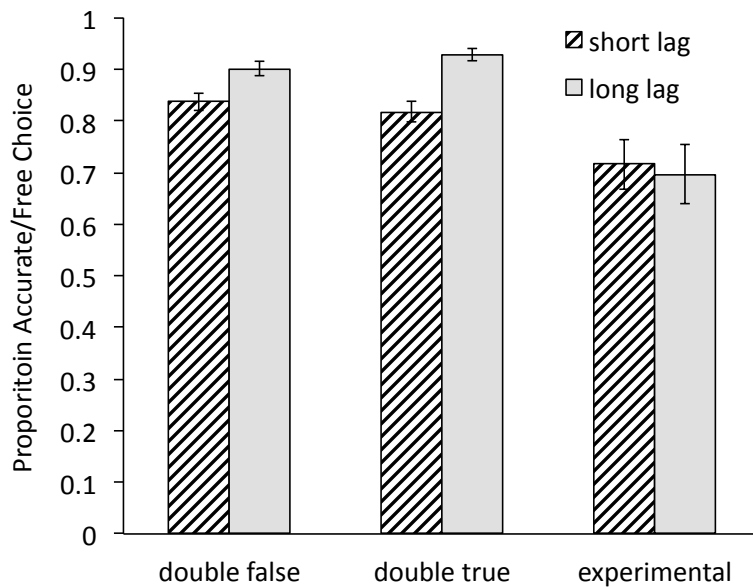
The choice proportions are shown in Figure 4. Accuracy dropped from the long lag to the short lag for true control sentences,  $M_l = 0.90$  ( $SD = 0.087$ ) vs  $M_s = 0.84$  ( $SD = 0.11$ ),  $\chi^2(1) = 4.28$ ,  $p = 0.039$ ,  $\beta = 0.56$ ,  $SE = 0.25$ ,  $t = 2.22$ , and false control sentences  $M_l = 0.93$  ( $SD = 0.079$ ) vs  $M_s = 0.82$  ( $SD = 0.12$ ),  $\chi^2(1) = 20.39$ ,  $p < 0.001$ ,  $\beta = 1.33$ ,  $SE = 0.25$ ,  $t = 5.26$ , but not for the experimental sentences,  $M_l = 0.70$  ( $SD = 0.37$ ) vs  $M_s = 0.72$  ( $SD = 0.31$ ),  $\chi^2 < 1$ , contrary to the scalar account of free choice inference.

The scalar implicature explanation for free choice inferences predicts that the rate of free choice responding should fall from the long lag to the short lag (not remain constant or increase). More precisely, Bott & Noveck (2004) observed that for scalar implicatures, the proportion of inference responses decreased from 0.44 in the long lag to 0.28 in the short lag. If free choice is indeed a complex scalar implicature, we should have observed an effect at least as large as Bott & Noveck and in the same direction. To test whether our effect (+0.02) was significantly different to that of Bott and Noveck (-0.16), we examined confidence intervals around the difference between the short and long lag means in our experimental sentences. This revealed that Bott & Noveck's effect,  $\text{logit}(0.16) = 1.66$ , fell outside our confidence intervals:  $M = 0.0015$ , 95% CI (2-tailed), [-1.26, 1.27] (using the standard error around the parameter estimates and assuming a normal approximation). We can thus be confident that any likely effect of the lag on the experimental sentences is less than the effect observed by Bott and Noveck, contrary to the free choice as complex scalar implicature hypothesis.

## 5.3 Discussion

If free choice inferences behave like scalar implicatures, the rate of free choice responding should decrease when participants are given a restricted processing time (following Bott & Noveck, 2004). We did not observe this pattern of results: the proportion of free choice inferences stayed more or less constant across the short and long lag conditions, whereas accuracy dropped significantly in the control conditions. Our data therefore provide further evidence against the scalar implicature account of free choice.





**Figure 4.** Choice proportions in Experiment 4. Double false and double true correspond to the control conditions. Error bars are the standard error of the mean.

## 6. General Discussion

According to recent theories (see section 1.3), free choice inferences are best analyzed as an extreme case of scalar implicatures. Consequently, the standard cost found for deriving a scalar implicature should be amplified. Yet, we found evidence for a reverse cost: *not* deriving a free choice inference is a costly phenomenon. This result argues against the otherwise parsimonious and widespread views that offer a single analysis for the two phenomena. We next discuss issues related to the methodology, before turning to the source of our effect and its implications for free choice theories.

### 6.1 Sentence verification tasks

We used a sentence verification task to measure full sentence reading times to free choice and literal interpretations. As with any methodology, there are some limitations to the conclusions that can be drawn. Here we discuss what those might be.

One potential criticism of the sentence verification task is that unconstrained verification times, such as we used, do not separate speed from accuracy with sufficient clarity to conclude that we were measuring speed of processing alone. For example, if there were a lower probability of retrieving the correct verification response for literal interpretations, participants may have delayed responding in order to maximize the accuracy (e.g., McElree & Nordlie, 1999, show that responses to metaphors might be delayed for this reason). A speed-accuracy trade-off (SAT) explanation seems unlikely to explain all of our results, however. First, participants would have had to have opposing response criteria and biases across our free choice tasks compared to our scalar implicature task or that of Bott and Noveck in order to explain the opposing pattern of results. Moreover, Bott et al. (2012) showed that the effects observed by Bott and Noveck could not be explained using SAT strategies. Since the two tasks were similar in many ways, it seems unlikely that SAT strategies could explain the difference we observe between free choice inferences and scalar implicatures. Second, we can see no reason why there would be a lower probability of correctly retrieving the literal

interpretation; the informativeness and verification requirements are in fact lower in the literal interpretation than in the free choice interpretation, exactly as for scalar implicatures.

A further difficulty with sentence verification tasks is that it is sometimes difficult to know why participants reject sentences. Perhaps participants were not rejecting the experimental sentences because they derived the free choice interpretation, but for some other reason. In which case the comparison between true and false was not a comparison between literal and free choice interpretations, as we have maintained, but between literal and general rejection. One possibility is that participants were rejecting statements because they did not make sense. For example, they may have been able to see both interpretations of the sentence and therefore rejected it because it was ambiguous. There are several arguments against this however. First, rejected sentences (false) interpretations were never slower than true interpretations whereas judging ambiguous sentences (and therefore responding false) might be expected to take longer than judging normal, or non-ambiguous, sentences. Second, accuracy to the experimental sentences in the free choice condition was generally very high ( $M > 0.93$  across both experiments). It seems unlikely that participants felt the sentences were ambiguous by the end of the training phase, given that they performed at such high levels of accuracy. Alternatively, one may argue that participants felt that the target sentences are ambiguous and recalled from the training phase that they should say false/true to ambiguous sentences. However, this possibility would not explain the current results and in particular the inverse mapping between false/true responses and short/long response times for free choice and scalar implicatures. Finally, if participants were rejecting sentences because they were underinformative or ambiguous, similar results would be expected for scalar implicature studies (underinformative sentences were rejected because they were infelicitous, say); yet we found the reverse pattern.

Finally, sentence verification tasks that use materials with a restricted range of sentence frames and target words, such as those in our study, are prone to participants noticing regularities in the sentences and developing nonlinguistic strategies. While we cannot eliminate the possibility of participants using nonlinguistic strategies, we note several factors that make it less likely that they were doing so: (1) we obtained the same pattern of results across a range of paradigms, that is, participants freely choosing the *or* interpretation (Experiment 1), being trained on one or the other interpretation (Experiment 2), and responding under restricted response times (Experiment 4). The results cannot therefore be explained by strategies being invoked by particular methods, e.g., feedback; (2) our experiments all involved a cover story, rather than general knowledge, which made the task relatively taxing and thus more difficult to form the abstract knowledge necessary for a strategy (c.f. Bott and Noveck, 2004); and (3) we cannot identify a strategy that would enable participants to perform accurately on the experimental sentences and the control sentences (participants who scored badly on the control sentences were removed), whilst also explaining our processing results. For example, it was suggested to us that participants with a free choice interpretation may be looking for false disjuncts and saying false as soon as they find one, while participants with a literal interpretation may be looking for a true disjunct and saying true as soon as they find one. While being consistent with the observed accuracy on controls and experimental sentences (indeed, the strategy is indistinguishable from one based on the meaning of *or*), the strategy fails to explain why we observed response time interactions between true and false control sentences and true and false experimental sentences.

In summary, while there are difficulties with sentence verification tasks, the range of techniques we have applied and the comparison between scalar implicatures and free choice have protected us from the most serious issues. More generally, we hope that more sophisticated techniques will be applied in the future to confirm and extend our findings, just

as has been the case in scalar implicatures (following the simple sentence verification task used by Bott & Noveck, 2004, many other researchers used other techniques: Breheny et al., 2005, used self-paced reading; Bott et al., 2012, used SAT; Huang & Snedeker, 2009, used visual world; Shetreet, Chierchia & Gaab, 2013, used fMRI; Tomlinson et al., 2013, used mouse-tracking).

## 6.2 Extending a paradigm: Truth value judgments and world-knowledge

Throughout this paper, we have made heavy use of the paradigm proposed by Bott and Noveck (2001) to study scalar implicatures. Their paradigm can be used to study the processing aspects of mechanisms that create alternative readings for a given sentence, just like scalar implicature enrichment mechanisms create new interpretations. The design is based on situations such that truth value judgments could serve as a signal for which of two readings has been derived. In their original study, Bott and Noveck used sentences such that world knowledge would make the two target readings have different truth values. But this manipulation constrained the sentences one can test. By playing with contextual information rather than relying on world knowledge we relaxed these constraints and showed that the same paradigm could be applied to study free choice inferences, and should be applied to study more phenomena (see Chemla and Bott, 2013 for an application to presuppositions).

In fact, the constraints imposed by world knowledge were problematic for scalar implicatures themselves, because the constraints on the stimuli led to the construction of odd sentences (e.g., *some elephants are mammals*). In Experiment 3, we used our contextual information manipulation to study scalar implicatures, and were able to replicate the well-known results about scalar implicatures with sentences which did not suffer from the same type of oddness (e.g., *some eagles were saved*).

## 6.3 Understanding the cost

In this section we discuss why we observed a cost to free choice responding. We make one precautionary note with respect to the interpretation of our study however. The study was designed to investigate the comparison between scalar implicatures and free choice interpretations, and not to investigate free choice *per se*. While our data categorically demonstrates that processing free choice is not like processing scalar implicatures, the data concerning whether free choice *or* is faster than literal *or* is more debatable. Across all four experiments we have shown that the size of any possible free choice cost is small relative to that predicted by scalar implicature account, which evidences our principle argument about the comparison, but whether significantly faster free choice responses are observed seems to depend on idiosyncratic factors (e.g., proportion of true and false responses). We feel that our data is highly suggestive of faster free choice than literal *or*, and this is how we continue the discussion, but it is possible that there is a small (and so far non-detectible) cost of free choice interpretations.

A faster free choice interpretation can be explained in two ways. First, free choice could be accessed before the literal meaning. (We are not assuming that the literal meaning is not computed along the way, but only that it is not a viable final option before free choice has been entertained). Perhaps the free choice interpretation is always derived and the processor sometimes enriches (or rejects) it and proceeds to derive the literal interpretation. Under this account, there is a cost associated with the literal interpretation because there is an extra processing stage involved in deriving the literal meaning.

The alternative is that readings are not accessed in this serial fashion. Specifically, it may be that one of them is accessed on each occasion, probabilistically or based on independent contextual considerations (as a constraint-based model might predict, e.g., MacDonald,

Pearlmutter, Seidenberg, 1994). The kind of effect we found could then emerge because the evaluation of one reading would be a more complex task than the same evaluation of the other reading. In our case, it may be faster to assess the truth value of the free choice reading. However, we see no independent reason for this, for example there is no sense in which free choice contains less information than non free choice. The formal relation between the two readings is the same as for scalar implicatures (the literal meaning in both cases is entailed by the other reading); yet the effect is the reverse. Similarly, one might argue the two meanings are accessed at the same time and the faster response times for free choice inferences are observed because there is a stronger preference for free choice overall. However, in Experiments 2 and 3, participants were provided with a context (training) to ensure that preference towards the literal and free choice interpretations respectively was approximately the same (>90% accurate responses in both conditions), yet we observed faster response times to the free choice interpretation. Furthermore, Experiment 4 demonstrated that after 1000ms there was an equally strong preference (70%) for free choice (inference) as there was for literal scalar implicature interpretations (72% in Bott & Noveck, 2004, Experiment 4). The time course of the emerging preferences therefore seems to be reversed for the free choice sentences relative to the scalar implicatures, just as with the reaction time analysis.

We are not in a position to make any stronger claims about the source of the free choice effect. In fact, despite the numbers of methodologies that have been used to study the scalar implicature effect, it is still debated what the source of the effect is (see Bott et al. 2012, Tomlinson et al., 2013, and Degen & Tanenhaus, under review, for a discussion of the sources of processing delay, and Katsos & Bishop, 2011, for discussions of the developmental delay). We hope that further techniques will be applied to get a finer-grained description of the processing pattern. What is important for our conclusion is the comparison between free choice and scalar implicatures and not on the individual explanations of their time course.

## 6.4 Implications for theories of Free Choice

The discrepancy between free choice and scalar implicatures argues against the otherwise parsimonious and widespread views that offer a single analysis for the two phenomena. In fact, they may favor views of free choice that require a reanalysis of more basic linguistic units such as disjunctions, conjunctions, epistemic and deontic modals, existential quantifiers, etc. (see section 1.3). Investigating the consequences for each of these alternatives would require cashing out the predictions they may offer about the relative order of the literal and free choice readings. It is not sufficient for an account to say that free choice inferences are part of the literal meaning, any account must explain how both readings are derived, and that the free choice reading should not come with a delay.

Instead of exhaustively reviewing existing formal accounts of free choice and confront each with our processing results (quite a challenging task in general), we would like to focus our discussion on scalar implicature accounts of free choice, which are our main target for these studies. This will lead us to discuss refinements of previous interpretations of other experimental results, and to finer distinction between recent approaches.

One may consider that the source of the cost to derive a scalar implicature corresponds to the relative (un)availability of alternatives. Specifically, one may need time to compute a scalar implicature not because the underlying Gricean-like reasoning requires time, but because it requires time to come to consider the relevant alternatives. One indirect argument for such a view comes from acquisition data. Children have been found to have difficulties with SIs, but these difficulties are reduced when alternatives are given to them explicitly (Gualmini et al., 2001). If that is the case, it is important to notice that the alternatives needed to derive free choice inferences are not only no more complex than the utterance (c.f. *some*

and *all*), but they are actually *simpler* than the utterance; in fact, they are contained in the original utterance (see Katzir, 2007 for considerations about how complexity enters in the determination of alternatives). Schematically, we are considering sentences of the form “A or B”, and the two relevant alternatives are simply the A part and the B part (see 9a-b). This is very different from regular scalar implicatures for which a lexical item is replaced by another. From this perspective, one may argue that scalar implicatures could give rise to a delay only if the alternatives involved are complex enough, i.e. in the *some* case but not in our free choice situations. This discussion reveals that, as of now, it is not decided what is the source of the time course we observe for scalar implicatures, let alone for other less studied pragmatic inferences such as free choice.

On the theoretical side, it is possible to capitalize on the similarities between free choice and scalar implicatures without equating them on processing grounds. More precisely, we have presented a view according to which free choice is an extreme case of a scalar implicature (a “double” scalar implicature, in a way, see Fox, 2007, and Franke, 2011, most explicitly). Building on the resemblance between the two phenomena (their cancelability, their properties under embedding, etc.), one may rather build a theory in which scalar implicatures are an extreme case of free choice, or in which the two sisters share a common root, but end up lying at the end of different branches. Although it would be out of place to develop this theory here, it is worth noting that Chemla (2008, 2010) is an instance of the latter kind of theory. In this approach, the common root of scalar implicatures and free choice is some abstract reasoning about alternatives. But the structure of the alternatives in the two cases (either simply “all” for the “some” scalar implicatures, or two parallel disjuncts “A” and “B”, in the case of free choice for disjunctions “A or B”) leads to quite different computations. In a nutshell, for scalar implicatures, the alternative (“all”) is negated, while for free choice, the two disjuncts are understood to have the same truth-value, and it does not rely on second-order implicatures. If it can be shown that the free choice type of computation is simpler than the “some/all” type, or if there is something to the discussion above about the role of the availability and complexity of the alternatives, then some unified approaches to scalar implicatures may remain serious contenders.<sup>10</sup>

## 7. Conclusion

Natural languages offer several means to manipulate and convey information. Accordingly, linguists and philosophers have described and classified diverse types of expressions and constructions that give rise to inferences that differ in nature. This classification is a representation of the underlying mechanisms that must be postulated to derive such and such inference (e.g., grammatical rules of composition, Gricean maxims). The resulting typology has been mostly defended on the basis of “offline” data (e.g., the plausibility of such and such discourse). We showed however that prominent theories lead to natural *processing* predictions. That is the case for instance for so-called free choice inferences. Free choice is of crucial importance to understand the way basic operations like disjunction and conjunction work in natural languages, because it challenges mainstream (and natural) hypotheses about these operators. As a solution to this puzzle, free choice is prominently assimilated to scalar implicatures. The consequence is that free choice should behave like scalar implicatures on processing grounds (as well as on offline data). We have

---

<sup>10</sup> Interestingly, the same processing pattern that we report for free choice has been found for presuppositions (Chemla & Bott, 2011). It is worth mentioning that Chemla (2008, 2010) offers a theory of presupposition as well, and the results align well with the picture that emerges from processing results: free choice and presuppositions rely on alternatives that are strictly simpler than the original utterance, while the alternatives involved in regular scalar implicatures are of equal complexity as the original utterance.

tested and invalidated this prediction. At the very least, our results place constraints on the adequate theory of free choice, both when it comes to formal modeling and to processing implementation.

Methodologically, we showed how to use processing data to inform linguistic classification enterprises. It requires both a careful examination of modern linguistic theories and a realization of their predictions at a cognitively relevant level. Unfortunately, specific cognitive implementations of formal theories are often missing or highly underspecified. By targeting *comparisons* between different types of inferences, however, progress can be made on the classification task without being committed to distort the formal models or post-hoc cognitive reconstructions of them.

## Acknowledgments

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

## References

- Alonso-Ovalle, L. (2008). Innocent exclusion in an alternative-semantics. *Natural Language Semantics* 16(2). 115–128. doi:10.1007/s11050-008-9027- 1.
- Baayen, R.H., Davidson, D.J. & Bates, D.M. (2008) Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59, 390-412.
- Bott, L., Bailey, T.M, 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., Frisson, S. & Murphy, G. L. (2008). Interpreting conjunctions. *Quarterly Journal of Experimental Psychology*, 62, 681-706.
- Bott, L., & Noveck, I.A. (2004) Some utterances are underinformative: The onset and time course of scalar inferences. *Journal of Memory and Language*, 51, 437–457.
- Breheny, R., Katsos, N., & Williams, J. (2006) Are generalised scalar implicatures generated by default? An on-line investigation into the role of context in generating pragmatic inferences. *Cognition*, 100,434–463.
- Chemla, E. (2008) Présuppositions et implicatures scalaires: études formelles et expérimentales. EHESS Dissertation.
- Chemla, E. (2010) Similarity: Towards a unified account of scalar implicatures, free choice permission and presupposition projection. Under revision for *Semantics and Pragmatics*.
- Chemla, E. & Bott, L. (2013) Processing presuppositions: dynamic semantics vs pragmatic enrichment. *Language and Cognitive Processes*, 28(3), 241-260.
- Chevallier, C., Noveck, I.A., Nazir, T., Bott, L., Lanzetti, V., Sperber, D. (2008). Making disjunctions exclusive. *Quarterly Journal of Experimental Psychology*, 61, 1751-60.
- Chierchia, G. (2004). Scalar implicatures, polarity phenomena, and the syntax/pragmatics interface. In Belletti, A., editor, *Structures and Beyond*. Oxford University Press.
- Chierchia, G., Fox, D., and Spector, B. (2008). The grammatical view of scalar implicatures and the relationship between semantics and pragmatics. In Maienborn, C., von Stechow, K., and Portner, P., editors, *An International Handbook of Natural Language Meaning*. Mouton de Gruyter.
- Degen, J. and Tanenhaus, M.K. (under review). Processing scalar implicature: A constraint-based approach.  
[http://www.bcs.rochester.edu/people/jdegen/docs/DegenTanenhaus\\_underreview.pdf](http://www.bcs.rochester.edu/people/jdegen/docs/DegenTanenhaus_underreview.pdf)

- Ducrot, O. (1969). Présupposés et sous-entendus. *Langue Française*, 4:30–43.
- Grodner, D., Klein, N. M., Carbary, K. M., & Tanenhaus, M. K. (2010). “Some”, and possibly all, scalar inferences are not delayed: Evidence for immediate pragmatic enrichment. *Cognition*, 116, 42–55.
- Fox, D. (2007). Free Choice and the theory of Scalar Implicatures. In Sauerland, U. and Stateva, P., editors, *Presupposition and Implicature in Compositional Semantics*, pages 537–586. New York, Palgrave Macmillan.
- Franke, M. (2011). Quantity implicatures, exhaustive interpretation, and rational conversation. *Semantics and Pragmatics*, 4(1):1–82.
- Geurts, B. (2005). Entertaining alternatives: Disjunctions as modals. *Natural Language Semantics* 13(4). 383–410. doi:10.1007/s11050-005-2052-4.
- Grice, H.P. (1975). Logic and conversation. *Syntax and Semantics*, 3(S 41):58.
- Gualmini, A., Crain, S., Meroni, L., Chierchia, G. & Guasti, M.T. (2001). At the semantics/pragmatics interface in child language. In *Proceedings of Semantics and Linguistic Theory XI*, Cornell, Ithaca: CLC Publications.
- Horn, L. R. (1989). A natural history of negation. University of Chicago Press Chicago.
- Huang, Y. & Snedeker, J. (2009) Online interpretation of scalar quantifiers: Insight into the semantics–pragmatics interface. *Cognitive Psychology*, 58, 376–415.
- Katzip, R. (2007). Structurally-defined alternatives. *Linguistics and Philosophy* 30(6). 669–690. doi:10.1007/s10988-008-9029-y.
- Kamp, H. (1973). Free choice permission. *Proceedings of the Aristotelian Society* 74. 57–74. <http://www.jstor.org/stable/4544849>.
- Kamp, H. (1978). Semantics versus pragmatics. In Franz Guenther & Siegfried Josef Schmidt (eds.), *Formal semantics and pragmatics for natural languages*, 255–287. Dordrecht: Reidel.
- Levinson, S. C. (2000). Presumptive meanings: The theory of generalized conversational implicature. MIT Press Cambridge, MA, USA.
- Katsos, N. & Bishop, D. V. M. (2011). Pragmatic tolerance: implications for the acquisition of informativeness and implicature. *Cognition*, 120: 67-81.
- Katsos, N. and Cummins, C. (2010). Pragmatics: From theory to experiment and back again. *Language and Linguistics Compass*, 4:282–295.
- Klinedinst, N. (2006). Plurality and possibility. Los Angeles: University of California, Los Angeles dissertation.
- Kratzer, A. and Shimoyama, J. (2002). Indeterminate pronouns: The view from Japanese. In Yukio Otsu (ed.), *Proceeding of the 3rd Tokyo conference on psycholinguistics*, 1–25.
- Lewis, D. (1979). A problem about permission. As reprinted in Lewis (2000): 20–33.
- Lewis, D. (2000). Papers in ethics and social philosophy. Cambridge studies in philosophy. Cambridge: Cambridge University Press.
- Magri, G. (2009). A theory of individual-level predicates based on blind mandatory scalar implicatures. *Natural Language Semantics*, 17(3), 245-297.
- MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). Lexical nature of syntactic ambiguity resolution. *Psychological Review*, 101 (4), 676 - 703.
- McElree, B. (1993). The locus of lexical preference effects in sentence comprehension: A time-course analysis. *Journal of Memory and Language*, 32, 536-571.
- McElree, B., & Nordlie, J. (1999). Literal and figurative interpretations are computed in equal time. *Psychonomic Bulletin & Review*, 6, 486-494.
- Giorgio Magri. A theory of individual level predicates based on blind mandatory scalar implicatures. *Natural Language Semantics*, 17:245–297, 2009.
- Noveck, I. A. and Posada, A. (2003). Characterizing the time course of an implicature: An evoked potentials study. *Brain and Language*, 85(2).

- Nieuwland, M. S., Ditman, T., & Kuperberg, G. R. (2010). On the incrementality of pragmatic processing: An ERP investigation of informativeness and pragmatic abilities. *Journal of Memory and Language*, 63, 324–346.
- van Rooij, R. and Schulz, K. (2004). Exhaustive Interpretation of Complex Sentences. *Journal of Logic, Language and Information*, 13(4):491–519.
- van Rooij, R. (2006). Free Choice Counterfactual Donkeys. *Journal of Semantics*, 23(4). 383–402.
- Sauerland, U. (2004). Scalar Implicatures in Complex Sentences. *Linguistics and Philosophy*, 27(3):367–391.
- Schulz, K. (2005). A pragmatic solution for the paradox of free choice permission. *Synthese* 147(2). 343–377. doi:10.1007/s11229-005-1353-y.
- Shetreet, E, Chierchia, C., & Gaab, N. (2013). When *some* is not *every*: Dissociating scalar implicature generation and mismatch. *Human Brain Mapping*. DOI: 10.1002/hbm.22269
- Simons, M. (2005). Dividing things up: The semantics of *or* and the modal/*or* interaction. *Natural Language Semantics* 13(3). 271–316.
- Spector, B. (2003). Scalar implicatures: Exhaustivity and Gricean reasoning. In ten Cate, B., editor, Proceedings of the Eighth ESSLLI Student Session, Vienna, Austria. Revised version in Spector (2007).
- Spector, B. (2007). Scalar implicatures: Exhaustivity and Gricean reasoning. In Aloni, M., Dekker, P., and Butler, A., editors, Questions in Dynamic Semantics, volume 17 of Current Research in the Semantics/Pragmatics Interface, pages 225–249. Elsevier.
- Sperber, D. and Wilson, D. (1995). *Relevance: Communication and cognition*. Oxford: Blackwell; Cambridge, Mass.: Harvard U. P.
- Tomlinson, J. M., Bailey, T. M. & Bott, L. (2013). Possibly all of that and then some: Scalar implicatures are understood in two steps. *Journal of Memory and Language*, 89(1), 18–35. doi: 10.1016/j.jml.2013.02.003.
- Veltman, F. (1996). Defaults in update semantics. *Journal of Philosophical Logic*, 25:221–261.
- Zimmermann, T.E. (2000). Free choice disjunction and epistemic possibility. *Natural Language Semantics* 8(4). 255–290. doi:10.1023/A:1011255819284.