frontiers

# Processing presuppositions and implicatures: Similarities and differences

**Cory Bill** [1], **Jacopo Romoli** [2,*] **and Florian Schwarz** [3]

[1] *ARC Centre of Excellence in Cognition and its Disorders, Department of Cognitive Science, Macquarie University, Sydney, NSW, Australia*
[2] *Ulster University, School of Communication and Media, Belfast, United Kingdom*
[3] *University of Pennsylvania, Linguistics Department, Philadelphia, Pennsylvania, United States*

Correspondence*:
Corresponding Author
j.romoli@ulster.ac.uk

## ABSTRACT

Presuppositions and scalar implicatures are traditionally considered to be distinct phenomena, but recent accounts analyze (at least some of) the former as the latter. All else being equal, this 'scalar implicature approach to presuppositions' predicts uniform behavior for the two types of inferences. Initial experimental studies comparing them yielded conflicting results. While some found a difference in the Response Time (RT) patterns of scalar implicatures and presuppositions, others found them to be uniform. We argue that the difference in outcomes is attributable to a difference in the type of response being measured: RTs associated with acceptance and rejection responses seem to pattern in opposite ways. Next, we report on a series of experiments to support this, and to compare the behavior of the two inferences more comprehensively. Experiments Ia and Ib look at both acceptance and rejection responses for both inference types, and find uniform patterns once the acceptance vs. rejection variable is factored in. Experiment II adds a new dimension by testing for the influence of prosody on the two inference types, and in this regard a clear difference between them emerges, posing a first substantive challenge to the scalar implicature approach to presuppositions. A third set of experiments investigates yet another prediction of this approach, according to which the presuppositional inference is introduced as a simple entailment in affirmative contexts. This predicts that these presuppositional inferences behave parallel to other entailments. Experiment IIIa compares rejections of affirmative sentences based on either their presuppositional inference or their entailed content and finds that they differ, with greater RTs for the former. As an additional control, Experiments IIIb and IIIc test for parallel differences between two entailments associated with *always*, which yield uniform results. In sum, while Experiments Ia and Ib are in line with previous findings that presuppositions and scalar implicatures under negation show uniform response time patterns, the differences found in Experiments II and IIIa-c pose a substantial challenge to approaches assimilating the two phenomena, while being entirely in line with the traditional perspective of seeing them as distinct.

**Keywords: Scalar Implicature, Presupposition, Inference, Processing, Semantics, Pragmatics**

## 1 INTRODUCTION

This paper experimentally compares two central linguistic inference types, namely Presuppositions (Ps) and Scalar Implicatures (SIs). Traditional approaches treat these as entirely distinct categories (Heim 1982; van der Sandt 1992; Beaver 2001 among many others). But recent approaches, building on a line of work going back to Gazdar 1979 and Wilson 1975 (among others), analyze at least certain presuppositions as scalar implicatures, largely motivated by the need to account for varying behavior of different presupposition triggers (Abrusán 2011; Abusch 2002, 2010; Simons 2001; Chemla 2009, 2010; Romoli 2012, 2015).[1] We begin with a sketch of the simplest possible form of this overall approach, directly assimilating scalar implicatures and presuppositions, which we refer to as the'SI approach to Ps,' and whose two core properties are schematized in (1-a) and (1-b).[2]

(1) **Properties**:
   a. In affirmative contexts, Ps are simply entailments.[3]
   b. In all other contexts (e.g., under negation), Ps are derived as SIs.

To illustrate (1-a), the presuppositional inference in (2-b) arising from (2-a), is a simple entailment according to this approach, just as (3-b) is an entailment of (3-a).[4]

(2) a. John stopped going to the movies.
   b. ⇝ *John used to go to the movies*

(3) a. John always went to the movies.
   b. ⇝ *John sometimes went to the movies*

Turning to the property in (1-b), the inference in (4-b), arising from the sentence in (4-a), is derived as an SI in contexts like negation, parallel to the derivation of (5-b) from (5-a).

(4) a. John didn't stop going to the movies.
   b. ⇝ *John used to go to the movies*

(5) a. John didn't always go to the movies.
   b. ⇝ *John sometimes went to the movies*

Two predictions that follow from the properties above are (6-a) and (6-b):

(6)

[1] Note that such approaches commonly differentiate between different types of presupposition triggers, and only propose to treat the inferences of a sub-class of traditional presupposition triggers as implicatures. Given our focus on triggers in the relevant sub-class, we simply refer to them as Ps here.

[2] Many of the proposals in the literature mentioned above depart from this simple version of the approach to some extent, by re-introducing some elements of difference between implicatures and presuppositions. These elements might affect the predictions in relation to the properties in (1-a) and (1-b) in different ways. We think that it is nonetheless useful to test experimentally the prediction of the simplest and most ambitious version of the approach and then take the results of that as a quantitative base to evaluate if and where a departure is needed from simply assimilating scalar implicatures and presuppositions. Recent pragmatic accounts to presuppositions like that in Schlenker 2008 also derive them in terms of conversational reasoning, though not equating them with scalar implicatures. This type of account makes non-trivial predictions in relation to the processing of presuppositions. Despite this distinction, we group it with the 'traditional approach' here and leave explorations of these predictions for further research.

[3] Traditional accounts are compatible with the assumption that presuppositional inferences in affirmative contexts are entailments, in addition to being presupposed, though this isn't necessarily extended to all presupposition triggers (see Sudo 2012 for discussion).

[4] The entailment from (3-a) to (3-b) actually involves some complications: in order for it to go through one has to assume that the restrictor of the universal quantifier *always* is non-empty. We leave this aside here, as it is orthogonal to our purposes; for discussion see Heim and Kratzer 1998, chapter 6.

**Predictions**: All else being equal,

a. in affirmative contexts, Ps and entailments should display uniform behavior.
b. in all other contexts, Ps and SIs should display uniform behavior.

We tested these predictions by comparing Ps to simple entailments, on the one hand, and to SIs, on the other. Specifically, we focus on the predictions in (6), in order to answer the question in (7). A positive answer to this question would be challenging for a unified approach to SIs and Ps, at least in its simplest version.[5]

(7) **Main question**: Do behavior patterns yield evidence for a distinction between Ps and entailments in affirmative contexts and between Ps and SIs in other contexts?

Previous studies in the literature have focused on the prediction in (6-b), comparing SIs and Ps directly, and have produced results that run against this prediction, based on delays in RTs found for SIs (Bott and Noveck 2004 and much subsequent work) on the one hand, and recent reports of the opposite pattern for Ps (Chemla and Bott 2013). We begin our discussion below with a review of these findings and contrast them with some other recent results reported by Romoli and Schwarz (2015), which found uniform RT patterns for Ps and SIs. We then argue, following a similar point made by Cremers and Chemla (2014), that the source of the difference in the results on Ps could well be due to a confound, namely a difference in terms of the types of responses — acceptances vs. rejections — being measured.

This motivates the first series of experiments reported here, which further extend the comparison between SIs and Ps. The results from Experiments Ia and Ib reconcile the conflicts between previous findings and show that once we look systematically at both acceptance and rejection responses, the evidence for a difference between Ps and SIs in RTs disappears. Thus, comparisons of RT patterns of the sort first employed in the study of SIs, testing the prediction in (6-b), do not challenge the SI approach to Ps. However, Experiment II clearly differentiates the two inference types by looking at the impact of prosodic stress on the inference-triggering expressions, which yields opposite effects for SIs and Ps. This poses a first challenge to the SI approach to Ps. An additional finding from our response time studies is that we do not replicate the previously reported general delays associated with SIs (e.g., Bott and Noveck 2004).

We then shift our attention to the prediction in (6-a) and report a third series of experiments that follow an approach presented in Kim (2007) and Schwarz (2016b). That is, these experiments look at rejections of sentences based on either their presuppositional inferences or their entailments. We find longer RTs for the former, which runs against the prediction in (6-a) and poses a second challenge to the SI approach to Ps.

[5] Let us emphasize here the 'all else being equal' element of these predictions. That is, these predictions are only claimed to apply in situations where the properties of the relevant meanings are as close to each other as possible. This is important as it increases the likelihood that any difference in the behavior patterns of the inferences is genuinely a result of the inferences being of different types. In line with this, we compared triggers that are as similar to each other as possible. Moreover, we would note that in our experiments the nature of the *uniformity* predicted in (6-a) and (6-b) varies somewhat depending on how close the situation is to the ideal of *all else being equal*. For example, in Experiment Ia and Ib we compare the processing profiles of three inferences that, according to the SI approach to Ps, are all derived as SIs. Despite this common derivational mechanism, there are other dimensions on which the relevant triggers vary (e.g., presence of negation), as a result, we take the 'uniformity' predicted by this approach to hold at a fairly general level. Specifically, for these experiments we test the prediction that, for each trigger, there will be uniformity in the general processing pattern produced when comparing responses motivated by an inference-based interpretation to responses based on a literal interpretation. At the beginning of each experiment we identify and justify the degree of behavioral uniformity predicted by the SI approach to Ps for the situation under investigation. Finally, in connection to the qualifications above, we also should make make note of work on 'scalar diversity' in the implicature literature, which has found differences across different scalar terms (Van Tiel et al. 2016, among others). The differences that have been found so far have chiefly been in the realm of inference derivation rates, but it is in principle possible for there to be within-inference variation in regards to other aspects of behavior as well. Nonetheless, when considering the simple version of the SI approach to Ps, outlined above, the differences we do find between SIs and Ps are not readily explained by scalar diversity. We will return to this later when discussing one such result, which is generated by Experiment II.

In sum, the results of Experiment II and those of Experiment IIIa-c challenge the SI approach to Ps by revealing differences between them where this approach predicts uniform behavior. This is further corroborated by differences between SIs and Ps found in previous work on language acquisition and language disorders (Bill et al., 2016; Kennedy et al., 2014). The overall evidence, then, is not in line with the predictions of the SI approach to Ps, as outlined in (6-a) and (6-b).

The paper is organized as follows. In section 2, we present the theoretical background on SIs, Ps, and the SI approach to Ps. In section 3, we discuss previous work on the processing of SIs and Ps and in particular those results taken as evidence for a difference between Ps and SIs. In section 4, we report our new series of experiments and in section 5 we discuss their implications for our main question and the processing of SIs and Ps. Section 6 closes the paper with some general conclusions.

## 2 BACKGROUND

### 2.1 The phenomena

Ps and SIs are inferences associated with certain expressions that go beyond the core lexically encoded, truth-conditional meaning. (8) and (9), repeated from above, illustrate inferences that are traditionally analysed as Ps and SIs, respectively.

(8) a. John didn't stop going to the movies.
b. ⇝ *John used to go to the movies*

(9) a. John didn't always go to the movies.
b. ⇝ *John sometimes went to the movies*

We focus on cases like (8) and (9) in particular, as they are maximally parallel, at least on the surface, in involving negation. But we also consider more standard cases of SIs in affirmative sentences such as (10). Sometimes the SIs in (9) and that in (10) are distinguished terminologically as "indirect" and "direct" ones (Chierchia 2004), and we will adopt this terminology.[6]

(10) a. John sometimes went to the movies.
b. ⇝ *John didn't always go to the movies*

One shared property of all these inferences is that they are not obligatorily present. In other words, in addition to "inference readings" illustrated above, all these sentences can have a "no-inference" reading as well, where the inference is absent. Consider (11) as compared to (8): the felicity of the continuation illustrates that the inference that John used to go to the movies is not necessarily present. The same goes for (12) and (13) and their inferences that John sometimes went to the movies and that he didn't always go, respectively.

(11) John didn't stop going to the movies . . . he never went!

(12) John didn't always go to the movies . . . (in fact) he never went!

(13)

[6] Roughly, the distinction is as follows: a direct SI is an SI arising from a weak scalar term in an upward entailing context and an indirect SI is one arising from a strong scalar term in a downward entailing context, such as the scope of negation. As we will see below, this distinction is purely terminological, as all theories of SIs that we know of treat direct and indirect SIs in the same way.

John sometimes went to the movies . . . (in fact) he always went!

This property, of course, is not shared by all inferences: in the case of a regular entailment like (14-b) of the sentence in (14-a), any attempt to suspend the inference, as in (15), results in infelicity, and the sentence sounds contradictory.

(14) a. John and Mary went to the movies.
b. ⇝ *John went to the movies*

(15) John and Mary went to the movies . . . #(in fact) John didn't go!

In light of this property any theory of SIs and Ps, unified or not, requires an account of (i) how these inferences arise to account for the inference readings, while (ii) also allowing for no-inference readings. In the next section, we briefly sketch how traditional approaches handle this challenge for SIs and Ps.

## 2.2 The traditional approach

In sketching standard analyses of Ps and SIs, we focus on the traditional approach, but for present purposes any account, old or new, which treats presuppositions and scalar implicatures as different falls in same class as the traditional perspective.

### 2.2.1 Presuppositions

Considering Ps first: the traditional approach is to analyse them as definedness conditions on admissible conversational contexts for the sentence carrying the presupposition. The gist of the idea is that a sentence like (16-a) is only felicitous in a context in which the presupposition in (16-b) is already assumed to be mutually accepted by the discourse participants (Stalnaker 1974; Karttunen 1974; Heim 1982, 1983; see also Beaver and Geurts 2012; Schwarz 2015; Romoli and Sauerland 2015 for an introduction to presuppositions).

(16) a. John stopped going to the movies.
b. ⇝ *John used to go to the movies*

In addition, an account of the so called 'projection' behavior of presuppositions is needed to explain how the presupposition of a sentence like (16-a) appears to be "inherited" by more complex sentences containing (16-a) such as (17), repeated from above.

(17) John didn't stop going to the movies.

Note that (16-a) and its negation in (17) both have the same presupposition that John used to go to the movies; in the traditional terminology, the presupposition of (16-a) in (16-b) 'projects' from the scope of negation in (17). Projection is not limited to negation, but is a general pattern involving all sorts of complex embeddings. For instance, the presupposition of (16-a) is also inherited by conditional sentences containing (16-a) in their antecedent, as well as questions or modal embedding (16-a): all of (18)-(20) standardly give rise to the inference that John used to go to the movies. In contrast, none of them convey that John is not going to the movies now, as entailments are interpreted relative to the embedding operators.

(18) If John stopped going to the movies, he must have gone to the gym more regularly.

(19) Did John stop going to the movies?

(20) John might have stopped going to the movies.

There are various well-developed proposals for accounting for presupposition projection in traditional terms, but we will not review these here in any detail for reasons of space. What is crucial for us, as before, is that all of these accounts treat presuppositions in a way that is very different from their treatment of SIs.

Finally, notice that traditional approaches quite generally assume presuppositions to be conventionally encoded in the lexical entries of the relevant expressions. This means that sentences containing a presupposition trigger necessarily introduce the corresponding presupposition. In order to reconcile this with cases of apparent suspension of presuppositions, as in (21), a further mechanism is assumed, e.g. one that 'accommodates' the presupposition locally, which results in the absence of any contextual constraints at the sentence level (Heim 1983; see also von Fintel 2008). This gives rise to the meaning paraphrased in (22), which is compatible with the continuation of (21), asserting that John never went to the movies.

(21) John didn't stop going to the movies . . . he never went!

(22) It's not true that (John used to go to the movies and stopped)
($\approx$ Either John didn't use to go to the movies or he didn't stop)

### 2.2.2 Scalar Implicatures

The traditional approach to SIs, which sees them as distinct from Ps, goes back to Grice (1975) and Horn (1972). On this approach, SIs can be understood as arising from the hearer reasoning about the speaker's communicative intentions. Take the inference in (23-b) based on (23-a).

(23) a. John sometimes went to the movies.
b. $\rightsquigarrow$ *John didn't always go to the movies*

In brief, the idea is that the hearer reasons that the speaker said (23-a), rather than something else, and in particular the more informative sentence in (24). Assuming that (24) is relevant to the purposes of the conversation, and that speakers are assumed to be committed to conveying the most informative relevant information at their disposal, the hearer will infer that the speaker's reason for not saying (24) is that the speaker believes (24) to be false. Therefore, the hearer derives the inference (23-b).[7]

(24) John always went to the movies.

A parallel line of reasoning, can be used to derive the indirect SI in (25-b) from (25-a). The hearer reasons that the speaker said (25-a), rather than the relevant and more informative (26). Therefore, the hearer infers that (26) is false, i.e., (25-b).

) a.

[7] We are skipping over a variety of details and assumptions here. See Gamut 1991 for a precise discussion of all the assumptions needed here to derive this inference.

John didn't always go to the movies.
b. ⇝ *John sometimes went to the movies*

(26) John didn't sometimes go to the movies (≈ John never went to the movies)

This brief review of the traditional perspective on Ps and SIs, while glossing over many intricacies, will suffice for our purposes. We primarily wish to provide a sense of how Ps and SIs are traditionally analyzed in clearly distinct ways. We now turn to more recent accounts of these inferences, in particular the SI approach to Ps.

## 2.3 The scalar implicature approach to presuppositions

The scalar implicature approach to presuppositions generally attempts to assimilate (certain) presuppositions to implicatures. In particular, some of the accounts within this general approach treat the presupposition associated with verbs like 'stop' as scalar implicatures of a sort (Simons, 2001; Abusch, 2002, 2010; Chemla, 2010; Romoli, 2012, 2015). In this section, we briefly sketch the simplest version of this approach focusing on sentences like (27-a) and its associated inference in (27-b):

(27) a. John didn't stop going to the movies.
b. ⇝ *John used to go to the movies*

Recall that one of the main phenomena to be accounted for is how the presuppositional inference of 'stop' arises from both affirmative and negated sentences. As mentioned, the traditional explanation is that (28), by virtue of the lexical entry of 'stop', is associated with the presupposition in (27-b), which then projects from the scope of negation in (27-a).

(28) John stopped going to the movies.

The SI approach to Ps offers a rather different explanation. First, (27-b) is simply (and only) an entailment of (28) on this account. This is in line with the observation that (27-b) is a non-cancelable ingredient of the overall meaning of (28), as asserting (28) and negating (27-b) sounds contradictory.

(29) #John stopped going to the movies but in fact he never went.

Assuming that (27-b) is an entailment of (28) is neither novel nor surprising: many accounts of Ps in the traditional approach share the view that the presuppositional inference is entailed in affirmative contexts. What is novel in the SI approach to Ps is to argue that (27-b) is *only* an entailment of (28). Second, the fact that (27-b) is standardly inferred from negated sentences like (27-a) as well is derived as a scalar implicature in a fashion parallel to the reasoning above for standard SIs. In particular, the idea is that the speaker said (27-a) rather than the relevant and more informative sentence (30). Therefore, the hearer infers that the speaker believes the latter to be false, which is equivalent to (27-b).

(30) John didn't use to go to the movies.

If this approach is correct, then the inferences associated with soft triggers such as *stop* are simply entailments when occurring in affirmative contexts, but (indirect) SIs when occurring under negation,

leading to the two key predictions in (6-a) and (6-b) above. On this view, verbs like *stop* are completely parallel to strong scalar items like *always*, which give rise to parallel inferences in positive contexts and in the scope of negation.

## 3 THE PROCESSING OF SCALAR IMPLICATURES AND PRESUPPOSITIONS

In this section, we briefly review previous work on the processing of SIs and Ps, focusing in particular on RT experiments.[8]

### 3.1 The processing of SIs

In recent years, research on scalar implicatures has undergone what Chemla and Singh (2014) call an 'experimental turn.' In particular, investigations of their processing properties have played a central role in the overall theoretical discussion. Most studies have focused on direct SIs but some recent studies have started looking at indirect ones, too. In a seminal paper, Bott and Noveck (2004) argue that SIs are associated with a delay in RTs. They investigated sentences like (31-a) and their direct SI in (31-b), which directly conflicts with common knowledge (as in fact all elephants are mammals). Based on the inference reading of the sentence, (31-a) should thus be judged 'false.'[9] As discussed above, however, the sentence also has a no-inference (or 'literal') 'some and possibly all' reading, which is compatible with common knowledge, and thus should lead to a 'true' judgment.

(31) a. Some elephants are mammals.
b. $\rightsquigarrow$ *Not all elephants are mammals*

The logic of the design in Bott and Noveck (2004) then is as follows: since 'false' responses are indicative of inference interpretations and 'true' responses of no-inference interpretations, measuring RTs for both types of responses should shed light on the time course of the availability of the two interpretations.[10] Their main finding, schematically represented in (32) (with $>$ indicating greater RTs) is that false responses were slower than true responses. They interpret this delay as showing that the computation of scalar implicatures involves additional processing efforts that go beyond those involved in the computation of literal meaning.

(32) **Bott & Noveck on DSIs**
inference readings $>$ no-inference readings

One particularly relevant version of their general approach trains participants prior to the main task to respond according to one or the other possible interpretations of the sentence in question. They find that participants that were trained to respond based on the no-inference interpretation were generally faster than those trained on the inference interpretation. Parallel results have been obtained in various similar studies since (Bott, Bailey, and Grodner 2012, among others), and also for implicatures associated with disjunction (Chevallier, Noveck, Nazir, Bott, Lanzetti, and Sperber, 2008). Other methodologies, such as reading times

[8] This section is adapted from Schwarz et al. (2015).

[9] Notice that the sentence in (31-a) is generally found to be somewhat odd, as is generally the case when scalar implicatures conflict with common knowledge (Magri 2010). This feature of the design is however shown not to be important in work replicating the main result of Bott and Noveck (2004), like that of Bott et al. (2012).

[10] There is an obvious potential concern about general difference between the time course of true and false responses, which Bott & Noveck try to address through different variants of their basic design. We will return to this issue when introducing our own study below.

(Breheny, Katsos, and Williams, 2006) and visual world eye tracking (Huang and Snedeker 2009b and following work) have yielded comparable results as well.[11]

Cremers and Chemla (2014) extend Bott and Noveck's approach to indirect scalar implicatures by looking at sentences like (33-a), with the inference in (33-b), which is again incompatible with common knowledge.

(33) a. Not all elephants are reptiles
b. ⇝ *Some elephants are reptiles*

Overall, they argue their findings to be parallel to Bott and Noveck's results, in that training participants to respond based on an inference interpretation vs. a no-inference interpretation gives yields slower responses for responses based on inference-readings than those based on no-inference readings:

(34) **Cremers and Chemla on ISIs**
inference > no-inference.

Note, however, that Cremers and Chemla (2014) report two experiments, with prima facie conflicting results. In the first one, without training, they actually found opposite results for DSIs and ISIs, as participants' 'false' responses were faster than 'true' responses for ISIs. However, they argue that this outcome is the result of confounds in the materials. First, subjects may have calculated implicatures for controls as well, due to the specifics of the overall stimuli in the experiment. Secondly, DSIs and ISIs differ in whether they contain 'matching' or 'mismatching' animal names and categories (e.g., *elephant* paired with *mammals* and *reptiles* respectively). Their second experiment avoided these confounds and statistically controlled for effects of polarity and truth value, and yielded results in line with those for DSIs, leading to the interpretation of their overall results outlined above. We will return to some related issues when discussing the investigation of Ps by Chemla and Bott (2013) below.

In sum, Bott and Noveck found that 'false' responses based on inference readings for direct SIs were slower than 'true' responses based on no-inference interpretations. Similarly, Cremers and Chemla found that 'false' responses based on inference readings for indirect SI were slower in comparison to 'true' responses based on no-inference readings. These results are in line with the general uniformity for direct and indirect SIs assumed in the literature, and with the initial interpretation by Bott and Noveck that scalar implicatures are associated with a delay.

## 3.2 The processing of Ps

The processing of Ps has been studied less than that of SIs. However, a number of recent studies have begun to fill this gap, using various processing measures to investigate Ps (see Schwarz 2015, 2016a). In this section, we review two recent RT studies on Ps that are directly relevant for our purposes. The first, by Chemla and Bott (2013), uses the paradigm of Bott and Noveck (2004) to look at Ps under negation, and yields results that appear to be very different from those for SIs. The second, by Romoli and Schwarz (2015), compares Ps (under negation) and (indirect) SIs directly and finds uniform RT patterns. These two results appear to be in direct conflict with one another and thus suggest opposite answers to our main question about the relationship between Ps and SIs. We discuss a possible source of the difference in outcomes, which motivates the first set of experiments reported below.

[11] Although other researchers have found different results using visual world eye tracking, which suggest implicatures are immediately available (e.g., Grodner et al. 2010; Breheny et al. 2013; Foppolo and Marelli 2017).

### 3.2.1 Chemla and Bott 2013

Chemla and Bott (2013) adapts the paradigm from Bott and Noveck (2004) to investigate Ps. The logic is entirely parallel: subjects judge sentences like (35-a) with the factive verb 'realise' (or, in their first experiment, 'know'), which gives rise to the presupposition in (35-b). This presupposition conflicts with common knowledge, and therefore, the sentence in (35-a) is only true on a no-inference reading.

(35) a. Zoologists did not realize that Elephants are reptiles.
b. $\rightsquigarrow$ *Elephants are reptiles*

Comparing the RTs of True vs. False responses provides a measure of comparison between the inference readings and the no-inference readings. Prima facie, their results suggest the opposite pattern of that found for SIs by Bott and Noveck (2004): True responses were slower than false responses, i.e., inference readings were faster than no-inference readings:

(36) **Bott and Chemla on Ps**
inference readings $<$ no-inference readings

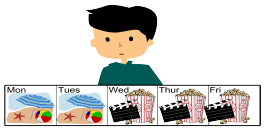
The interpretation proposed by Chemla and Bott (2013) is that the computation of P-inferences, unlike that of SI-inferences, does not incur a delay, suggesting that the inferences involved are different, at least in the way they are processed. This poses a challenge for the SI approach to Ps. Note however, that the confound from the first experiment by Cremers and Chemla (2014) arising for indirect SIs is relevant for the present results for Ps as well: recall that the indirect SI materials involved a mismatch with respect to the relationship between the name of the animals mentioned (e.g., *elephants* paired with *reptiles*), which the authors argue might have hindered acceptance of sentences like (33-a). Recall also, that for direct SIs, the relevant targets instead involve a match between name and category, so conversely this might have facilitated the acceptance of sentences like (37).

(37) Some elephants are mammals.

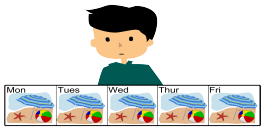
Turning back to the experiment in Chemla and Bott (2013), it is entirely parallel with the situation in Cremers and Chemla (2014). That is, unlike in Bott and Noveck, the target sentences in Chemla and Bott (2013), such as (35-a), involve a mismatch between the name and the category. As suggested by Cremers and Chemla (2014) for their own results, this factor could have influenced the results of Chemla and Bott (2013). That is, the increased RTs associated with no-inference readings could have been caused by this mismatch, rather than different derivational mechanisms. The existence of this potential confound means that the results in Chemla and Bott 2013 have to be interpreted with caution, and without implementing the same kinds of control techniques as Cremers and Chemla (2014) use in experiment 2, they do not conclusively establish any difference between SIs and Ps.

### 3.2.2 Romoli and Schwarz 2015

Recently, in a study by Romoli and Schwarz (2015) RTs for Ps and SIs under negation were directly compared to one another. In this study, instead of a direct truth-value judgment task, a version of a sentence picture matching task was used (Huang et al., 2013). This paradigm records both response choices and response times as dependent variables. A sentence was presented to participants and they were directed to pick a picture, from a set of three, that best matched the sentence. Each of the pictures depicted an




*Always*-Target (+LIT/+INF)




*Always*-Target (+LIT/-INF)

**Figure 1.** Target pictures for *always* conditions, matched with a sentence like (38).

individual and a 5-day calendar strip, with each day being filled with an iconic representation of an activity that the individual had engaged in on that day (see Figures 1 & 2). In addition to these two 'visible pictures' there was a 'Covered picture'. Participants were told that one of the three pictures was a match for the presented sentence. One of the visible pictures was a 'Target picture', which was either consistent or inconsistent with the inference ('+LIT/+INF' vs. '+LIT/-INF' condition).[12] The second visible picture was a distractor and so was incompatible with both possible interpretations. Participants were told that if neither of the visible pictures were a good match, then they should select the Covered picture.

(38) John didn't always go to the movies last week.

The +LIT/+INF Target picture depicts the character going to the movies on several days, making it consistent with the 'sometimes' implicature of 'not always'. In contrast, the +LIT/-INF Target picture depicts the character never going to the movies, making it inconsistent with this implicature. By comparing the RTs associated with Target choices in these two conditions Romoli and Schwarz (2015) were able to compare the processing of different interpretations based on the same type of response.[13]

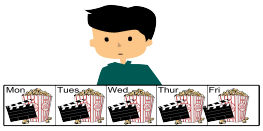
Similarly, for the *stop* condition, participants would evaluate sentences like (39) against one of the two overt pictures in Figure 2, a distractor picture and a Covered picture. Again the +LIT/+INF Target picture was compatible with the inference interpretation of the sentence, while the +LIT/-INF Target picture was only compatible with the no-inference interpretation.

(39) John didn't stop going to the movies on Wednesday.

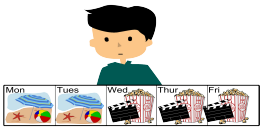
Unsurprisingly, the Target picture in the +LIT/+INF condition was chosen at ceiling level, while the +LIT/-INF condition yielded more mixed results. But most importantly, the RT results for Target choices were uniform for Ps and SIs, as schematized in (40), in that RTs in the +LIT/+INF conditions were significantly faster than in the +LIT/-INF conditions, in contrast with the findings discussed above. (Note that while the +LIT/+INF picture could be accepted on either a no-inference or an inference interpretation, the difference in RTs suggests that at least a sizable portion of Target choices was based on the latter; this assumption justifies the use of 'inference' and 'no-inference' in the schematic illustration below, and will also be utilized in the data analysis of the experiments in the next section.)

[12] Romoli and Schwarz (2015) label the conditions INFERENCE-TRUE and INFERENCE-FALSE respectively; we choose the more transparent labels here to clearly signal that the images shown in the former can in principle be accepted on either a literal or an inference interpretation.

[13] Note that, in principle, selection of the +LIT/+INF Target picture could also be motivated by a no-inference/literal interpretation. However, if all these selections were based on such an interpretation, then we would expect participants' behavior in these two conditions to be equivalent. Therefore, the fact that Romoli and Schwarz (2015) found substantial variance in the RT results, suggests that, at least a sizable portion of Target picture selections in the relevant condition are motivated by inference interpretations.




(i) *Stop*-Target (+LIT/+INF)




(ii) *Stop*-Target (+LIT/-INF)

**Figure 2.** Target pictures for *stop* conditions for a sentence like (39).

(40) a. **Romoli and Schwarz 2015 on indirect SIs**
inference < no-inference.
b. **Romoli and Schwarz 2015 on Ps**
inference < no-inference.

Note that the results for Ps here seem to be in-line with those in Chemla and Bott (2013), in that inference readings were faster than no-inference readings. The result for indirect SIs, however, is puzzling in that it appears to be exactly the opposite of what Cremers and Chemla (2014) find in their experiment 2. Moreover, with regards to our main question in (7), these results suggest that Ps and SIs (at least indirect ones) do not differ in their RT patterns after all, which would be consistent with a uniform account of SIs and Ps. This raises the question of what is behind these seemingly conflicting findings. One possibility relates to differences in the types of responses that were compared between these studies. As mentioned, previous response time studies generally explored the relevant inferences by comparing 'true' responses to 'false' responses. And, while Cremers and Chemla (2014) attempted to control for any effect of response-type, the more reliable way of controlling for such an effect is to compare the same kind of responses, which the setup of Romoli and Schwarz (2015) made possible. To put it another way, Romoli and Schwarz (2015) raise the possibility that the method employed by previous studies may have been undermined by a confound. Specifically that, rather than only being influenced by the interpretations of interest, participants' responses may have also been influenced by the nature of the response provided (i.e. sentence acceptance vs. rejection). The experiments reported in the next sections were designed to investigate this issue by further exploring the relationship between Ps and SIs.

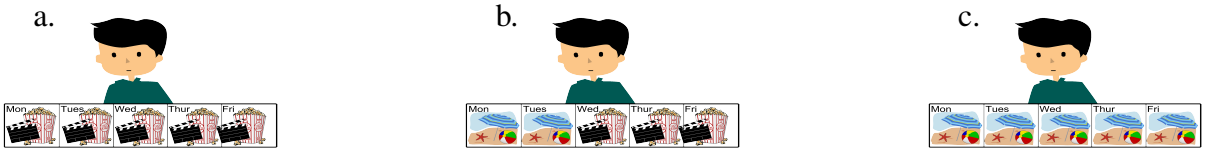
## 4 THE EXPERIMENTS

In this section, we report on three series of experiments testing the two predictions of the SI approach to Ps outlined in (6-a) and (6-b).

### 4.1 Experiment Ia

The first experiment adopted the approach taken in Romoli and Schwarz (2015) and applied it to investigating whether there are processing pattern differences between direct and indirect implicatures when we compare alike responses.[14] This allows for a more comprehensive comparison to the results from Bott and Noveck (2004) and Cremers and Chemla (2014) on the one hand, and Romoli and Schwarz (2015) on the other. It also offers a more comprehensive perspective on the role of response type in RT patterns. Note that, for this experiment (and Experiment Ib), the relevant uniformity prediction is that the relative

[14] This experiment was first reported in Schwarz et al. (2015), from which this subsection is adapted.




| | | | |
|---|---|---|---|
| +LIT/-INF | +LIT/+INF | -LIT/+INF | **(DSI)** |
| -LIT/+INF | +LIT/+INF | +LIT/-INF | **(ISI)** |

(i) John sometimes went to the movies ⇝ *John didn't always go* **(DSI)**

(ii) John didn't always go to the movies ⇝ *John sometimes went* **(ISI)**

**Figure 3.** Target Picture versions and conditions for Experiment 1.

processing patterns of each trigger will be similar. That is, the prediction is not that the RTs will be exactly the same as the relevant triggers differ substantially in other ways; namely, the presence of negation in one and not the other. Instead, the prediction is that the overall RT pattern, created by comparing inference and no-inference interpretations, will be similar. To gain a full comparison, we looked at both target choices (acceptance judgments) based on inference and no-inference interpretations, and Covered picture choices (rejection judgments) based on both types of interpretation.

### 4.1.1 Methods

#### *4.1.1.1 Materials & Design*

Following Romoli and Schwarz (2015), we used the Covered picture paradigm (Huang et al., 2013), with both response choices and RTs as dependent variables. Participants were presented with two pictures, one of which was simply black and was introduced as covering a hidden picture.[15] The instructions provided a detective scenario, where information about a suspect was presented as having been extracted from intercepted communication, and the participant's task was to decide which of two potential culprits fit the provided description. It was explicitly stated that only one of the two pictures would match the description, so that the Covered picture should only be chosen in situations where the overt picture did not match the sentence. We believed this setup would increase the chance of participants basing their responses on no-inference interpretations for the following reasons: First, the described source of the information remained opaque due to its nature of stemming from intercepted communication, which makes it uncertain whether the speaker of that sentence was fully informed. Secondly, the emphasis that only one picture would match the description provided by the sentence should increase target choices for +LIT/-INF pictures, on the assumption that no-inference interpretations are in principle available but generally somewhat dispreferred. That is, as the Covered picture could be completely 'False', if there is a possible reading that makes the Target picture 'True' the participant has a good reason to go with that reading, even if it is a dispreferred reading. At the same time, as noted above, having the Covered Picture as a response option ensures that subjects need not feel forced to give a response that they may feel uncomfortable about.

The basic logic of the design was parallel to that of Romoli and Schwarz (2015), in that the overt Target picture either was consistent with a given interpretation or not. More concretely, sentences (i) and (ii)

[15] Note that, unlike Romoli and Schwarz (2015), we didn't include a 'distractor picture'. This change was done merely to simplify the material and was not expected to have any substantive effect on the results.

in Figure 3 were displayed with one of the pictures in Figure 3 and a Covered picture.[16] For the DSI condition with *sometimes*, the picture in Figure 3a is only compatible with a no-inference interpretation, as the depicted person always went to the movies. Target choices in this case must therefore be based on the no-inference interpretation. Covered picture choices for this picture in turn are indicative of inference interpretations. The picture in Figure 3b is consistent with an inference interpretation (as well as a no-inference interpretation, since it is entailed by the inference interpretation), so target choices are generally expected here. Finally, the picture in Figure 3c is inconsistent with both interpretations, as the depicted individual never went to the movies, so Covered picture choices are expected here. For purposes of analysis, this design allowed us to compare Target and Covered picture responses to the picture in Figure 3a to Target and Covered picture responses in the control conditions in Figures 3b and 3c respectively. Thus, this set up provides a comparison between inference-based rejections (Covered picture choices for Figure 3a) and literal meaning based rejections (Covered picture choices for Figure 3c), as well as between no-inference acceptances (target choices for Figure 3a) and inference acceptances (target choices for Figure 3b, assuming as above that at least a sizable portion of responses here is based on an inference interpretation).

The same general logic applies to the ISI sentences (ii), though with different mappings onto the pictures. The picture in Figure 3c serves as a test for no-inference interpretations, as target choices are incompatible with the inference that John sometimes went to the movies. Covered picture choices for this pictures in turn must be based on inference interpretations. The picture in Figure 3b is consistent with the inference interpretation (as well as a no-inference interpretation, as for DSIs), and the picture in Figure 3a is inconsistent with either interpretation. So in the case of ISIs, Figure 3c is expected to yield a mix of target and Covered picture choices, depending on the interpretation participants base their judgments on in a given trial, which can be compared to the Covered picture and target choices in the respective control conditions.

Let us expand here on our assumption about the correspondence between responses and the interpretation that they are based on. As pointed out already, in certain conditions, it is not clear whether certain picture selection choices are motivated by an inference or a no-inference interpretation. Specifically, target choices for Figure 3b and Covered picture choices for Figure 3c could be based on either inference or no-inference interpretations. This is because both interpretations are consistent with Figure 3b and inconsistent with Figure 3c. However, if we assume consistency in participant's interpretations between conditions, then we can discern whether any of these responses are based on inference interpretations by comparing responses to Figures 3b & c to a condition without this ambiguity. For example, in the case of the DSIs condition, Figure 3a is only consistent with a no-inference interpretation. Therefore, if the participant group selects more covered pictures when presented with Figures like 3a than with Figures like 3c, then it is likely that at least some of the latter Covered picture selections were motivated by inference interpretations. Similarly, Target picture selections of Figure 3b can be compared with Target picture selections of 3a to determine if any of the former were motivated by no-inference interpretations. A similar comparison between conditions can be done in the ISI condition. (In addition to response patterns, differences in RTs also support this assumption, as noted already for Romoli and Schwarz (2015) above.)

[16] Note that the condition labels presented in Figure 3 relate to the truth-value of the two critical elements of the sentence; namely, the literal content and the inferential content. For example, in the case of the condition '+Lit/-Inf' for the DSI sentence, the picture is consistent with the literal content that *John went to the movies at least once*, but is inconsistent with the inference that *John didn't always go to the movies*. Moreover, in the case of the '-Lit/+Inf' conditions, the target picture should not be able to be selected, due to it not satisfying the literal content of the relevant sentence, despite the fact that it is consistent with the inference (corresponding to the literal meaning of the paraphrase).

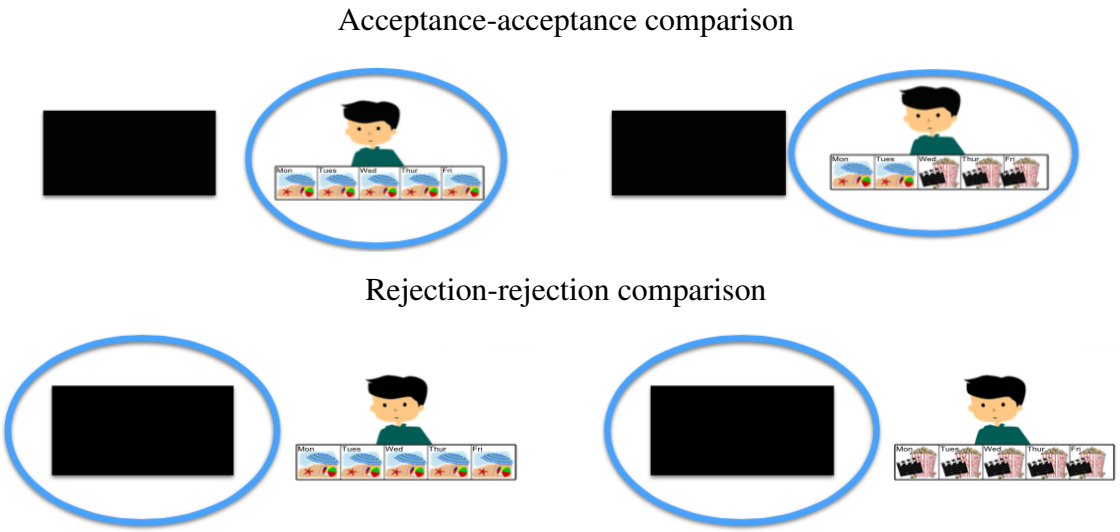



**Figure 4.** Acceptance-acceptance and rejection-rejection comparisons for ISI sentences

Figure 4 summarizes the two critical comparisons in the ISI conditions in the display format used in the experiment: no-inference acceptance vs inference acceptance ('acceptance-acceptance' comparison) and inference-rejection versus no-inference rejection ('rejection-rejection' comparison).

#### *4.1.1.2 Participants & Procedure*

35 undergraduate students from Macquarie University participated in the study. They saw 36 sentence picture pairs of the sort described above, with 6 items for each pairing, counterbalanced across participant groups. In addition, there were a total of 36 filler items; 18 were variants of the experimental items containing *always* without negation, paired with all three picture types to ensure that pictures such as those in Figures 3a/c were viable target choices throughout the experiment sufficiently often. There also were 6 items containing plain negation (e.g., *John didn't go to the movies last week.*), again paired with the various picture types to even out choices of types of pictures. Finally, 12 items were from another sub-experiment containing negation and *again*. At the beginning of the experiment, participants were presented with instructions laying out the detective scenario described above. They then were shown some example sentences and pictures, and completed a total of 4 practice trials (none of them resembling the crucial experimental conditions) to ensure they understood the Covered picture setup. Throughout this initial phase, they were free to ask any clarification questions. After this, presentation of the experimental trials began.

### 4.1.2 Results & Discussion

For purposes of statistical analysis, responses were coded according to whether they were based on their relation to an inference reading. Target selection of the pictures in Fig. 3a (DSI) and Fig. 3c (ISI) clearly indicates a no-inference reading, whereas Covered Picture selection for these pictures unambiguously reflected an inference reading. Accurate responses in the other conditions were compatible with both inference and no-inference readings, but were coded in terms of the strongest reading on which they could be based. For example, acceptance of the Target picture in 3b was coded as an inference response, though of course a positive instantiation of an inference reading entails truth of a no-inference reading as well. The negative response towards the Target picture for the versions in Fig. 3c (DSI) and 3a (ISI), as reflected in

| Inference Type | +LIT/-INF (Fig. 3a/c) | -LIT/+INF (Fig 3c/a) | +LIT/+INF (Fig 3b) |
|---|---|---|---|
| DSI | 22.9 | 0.005 | 97.1 |
| ISI | 50.9 | 0.005 | 95.7 |

**Table 1.** Target choice rates in % by condition

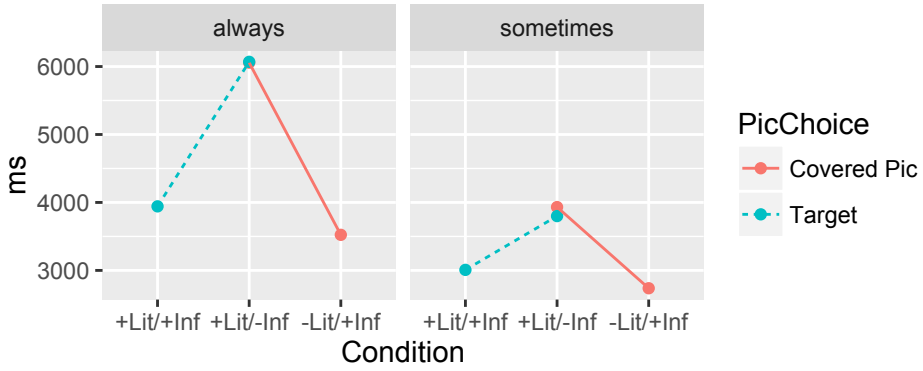



**Figure 5.** RTs for responses by picture choice and condition. +LIT/+INF target choices and +LIT/-INF Covered picture choices are taken to reflect inference interpretations, and +LIT/-INF target choices and -LIT/+INF Covered picture choices no-inference interpretations.

selection of the Covered Picture, was coded as a no-inference response, though again, a negative relation of a no-inference reading towards a picture entails a negative relation for the inference reading as well. This coding decision is not crucial for the overall interpretation of the data, but we think it reflects the difference across conditions in terms of whether the two readings are in conflict or not reasonably well. Target choice proportions as well as RTs (measured from the display of the sentence, which was added to the screen 800ms after the picture was first shown) were analyzed.

#### *4.1.2.1 Response rates*

Mean target selection rates are provided in Table 1. Accuracy in the conditions where both literal and inference interpretations led to the selection of the same image (Figures 3b/c for DSIs, Figures 3a/b for ISIs) were at ceiling, as expected. Both inference and no-inference (i.e. literal) interpretations occurred in the DSI and ISI +Lit/-Inf conditions, but inference interpretations occurred more often with DSIs than with ISIs, as there were fewer Target picture choices for DSIs. A planned comparison between these two conditions using a logistic regression mixed-effect model revealed this difference in implicature-response rates to be significant ($\beta = 4.01$, $SE = 0.98$, $z = 4.07$, $p < .001$).

Note also that the difference between the +LIT/+INF and +LIT/-INF responses suggests that at least some of the Target picture selections in the former condition were a result of participants accessing an inference interpretation. That is, if participants were only accessing literal interpretations for our test sentences, you would expect the response rates in these two conditions to be the same.[17]

[17] Similarly, the Covered picture selections between the -LIT/+INF and +LIT/-INF conditions suggests that some of these selections in the former condition were a result of accessing an inference interpretation.

#### *4.1.2.2 Response Times*

The mean RTs for all conditions are illustrated in Figure 5. Note that seeing this from the perspective of inference vs. no-inference interpretations as laid out above, yields a cross-over interaction pattern, showing that the relation between RTs for inference and no-inference interpretations depends crucially on whether we look at acceptances in the form of target choices or rejections in the form of Covered picture choices. In the former case, inference interpretations are faster than no-inference ones, while the reverse holds in the latter.

To investigate this result statistically, we analysed both the DSI and ISI subsets of data as a $2 \times 2$ interaction design with response (Target vs. Covered picture) and interpretation (inference vs. no-inference) as factors, using mixed-effect models with subjects and items as random effects, as implemented in the *lmer* function of the *lme4* package in $R$ (Bates, 2005). Following Barr et al. (2013), we used the maximal random effect structure that would converge, with random effect slopes for each factor, as well as the interaction, if possible. To assess whether inclusion of a given factor significantly improved the fit of the overall model, likelihood-ratio tests were performed that compared two minimally different models, one with the fixed effects factor in question and one without, while keeping the random effects structure identical (Barr et al., 2013). We report estimates, standard errors, and t-values for all models, as well as the $\chi^2$ and $p$-value from the likelihood-ratio test for individual factors. The statistical details are summarized in Table 2. The $2 \times 2$ interactions were highly significant for both ISIs and DSIs, as were the relevant simple effects comparing inference vs. no-inference responses by response type. Schematically, the results can be summarized as follows:

(41) RT patterns for Scalar Implicatures (for both DSIs and ISIs):

a. **rejection response**
inference $>$ no-inference

b. **acceptance response**
inference $<$ no-inference

The results for acceptances (Target-choices), where implicature-based responses were faster than those only compatible with the literal meaning, are entirely in line with the findings by Romoli and Schwarz (2015) for ISIs, but constitute a novel finding for DSIs. The finding that inference-based rejections (Covered Picture-choices) were slower for both types of implicatures prima facie seems to be in line with previous findings for DSIs from Bott and Noveck (2004) on, and with the findings by Cremers and Chemla (2014) for ISIs. However, note that the comparison we make is one between a condition where a Covered Picture choice can be unambiguously attributed to an inference interpretation (the equivalent of saying 'false' to *Some elephants are mammals.*), and a condition where the literal meaning suffices to lead to a Covered Picture choice, but an inference interpretation would have led to the same result (the equivalent of saying 'false' to *Some elephants are insects.* - B&N's control T3). Similarly, our acceptance comparison is between acceptances that are unambiguously based on a no-inference reading and ones where inference and no-inference readings yield the same result (parallel to B&N's T2 control: *Some mammals are elephants.*). The comparison within our data that is truly on par with the crucial comparison of Bott and Noveck (2004) (as well as Cremers and Chemla 2014) is the one between Covered Picture choices based on an inference interpretation and Target choices based on a no-inference interpretation. But here, we find no significant difference at all.

| DSI's | $\beta$ | $SE$ | $t$ | $\chi^2$ | $p$ |
|---|---|---|---|---|---|
| **Interaction** | 2119.1 | 563.4 | 3.76 | 9.67 | <**.01** |
| **Simple Effects:** | | | | | |
| Covered Picture Choices: inference > no-inference | -1418.6 | 534.8 | -2.65 | 6.38 | <**.05** |
| Target Choices: inference < no-inference | 666.1 | 276.5 | 2.41 | 5.42 | <**.05** |
| ISI's | | | | | |
| **Interaction** | 5902.7 | 1793.5 | 3.29 | 9.67 | <**.01** |
| **Simple Effects:** | | | | | |
| Covered Picture Choices: inference > no-inference | -3302.2 | 881.6 | -3.75 | 7.80 | <**.01** |
| Target Choices: inference < no-inference | 2197.9 | 580.2 | 3.788 | 11.734 | <**.001** |

**Table 2.** Summary of response time analyses: Interaction between Picture Choice and inference status and simple effects for relevant paired factor levels.

Now, let us consider these results in light of the SI approach to Ps' prediction of uniform processing patterns between DSIs, ISIs, and Ps, (i.e., (6-b)). Once we considered the acceptance versus rejection factor, DSIs and ISI exhibited uniform RT patterns, contrary to initial appearances from Romoli and Schwarz (2015). Next, we turn to Ps considered from the same, more comprehensive perspective, to see whether this uniformity might extend in the manner proposed by the SI approach to Ps.

## 4.2 Experiment Ib: *Stop* in negated sentences

In Experiment Ib, we used the same methods as in Experiment Ia to extend the investigation above to Ps, and in so doing, address the main question of this paper regarding the relationship between Ps and SIs. That is, to test the SI approach to Ps' prediction that the processing patterns of SIs and the relevant Ps should be uniform. Note that, as in Experiment Ia, the uniformity prediction that we are testing is the expectation that the relative processing patterns of Ps will be the same as SIs, not that the RTs will be exactly the same across these inferences.

### 4.2.1 Methods

#### *4.2.1.1 Materials & Design*

We used the same Covered picture paradigm as in Experiment Ia, with two pictures and both response choices and RTs as dependent variables. The basic logic of the design was also identical to that of Experiment Ia, but this time we were looking at presuppositional sentences. The stimuli included both sentences with and without negation. However, as laid out in the introduction, only the case of soft triggers under negation lends itself to a direct comparison with SIs (and specifically ISIs). We therefore focus the discussion in the present section on that case. The case of 'stop' in affirmative sentences will be discussed separately in Section 4.4. An illustration of the negative conditions is provided in Fig. 6. The sentence in Figure 6 was displayed with one of the pictures in Figure 6 and a Covered picture.

The picture in Figure 6a, paired with the negative 'stop' sentence, constitutes the Target-selection control, as both the putative presupposition (that John went to the movies before Wednesday) and the asserted part (that he went to the movies from Wednesday on) are true. The picture in Figure 6c provides the Covered Picture-selection control, as the asserted part is false (since he did stop going to the movies), although the presupposition is true. Figure 6b constitutes the critical case, as the putative presupposition is false, while the assertion is true. If a participant accesses an inference interpretation, the Covered Picture should be

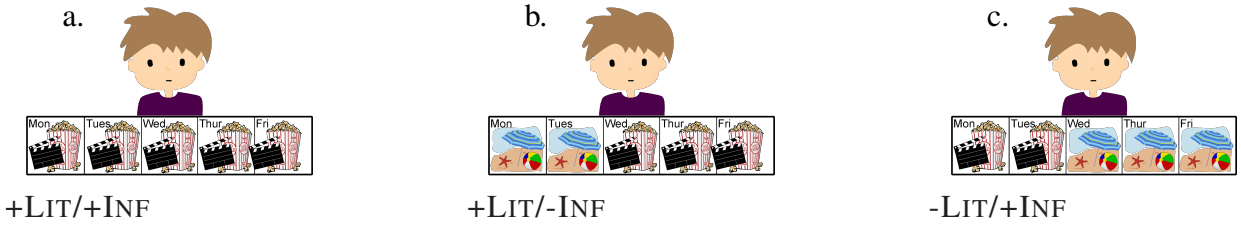



**Figure 6.** Target Picture versions and conditions

chosen. If a participant accesses a no-inference interpretation the Target picture should be selected. As in Experiment Ia, responses to Figure 6b were coded as inference and no-inference responses respectively, based on whether the Covered picture or the Target picture was selected. Figures 6a and c were taken to provide controls with the same response for the respective critical trials.

#### *4.2.1.2 Participants & Procedure*

34 undergraduate students from the University of Pennsylvania participated in this study for course credit. Each saw 6 sentences in the +LIT/-INF and 6 in the -LIT/+INF conditions, and these were drawn from a total of 24 sentences. The other 12 were shown in the affirmative condition (discussed below), and the condition in which a given item was shown was counterbalanced across four groups of subjects. Another 12 items were presented in the +LIT/+INF condition, again drawn from a total of 24, with counter-balancing between it and an affirmative variant. In addition, there were 21 fillers from another sub-experiment. Instructions and practice trials were as described for Experiment Ia.

### 4.2.2 Results & Discussion

#### *4.2.2.1 Response rates*

Unsurprisingly, the Target-selection rates for the control conditions were at ceiling and floor for the respective control conditions. In the critical condition, the Target was selected 62% of the time, which was significantly higher than in the -LIT/+INF control ($\beta = -4.63$, $SE = 0.82$, $z = -5.63$, $p < .001$), but also significantly lower than in the +LIT/+INF control ($\beta = 3.11$, $SE = 0.71$, $z = 4.38$, $p < .001$).

#### *4.2.2.2 Response times*

The RT results are summarized in Fig. 7. We find a pattern that is generally parallel to that for implicatures, and which corresopnds to a cross-over interaction between type of reading (inference vs no-inference) and type of response (acceptance vs rejection) when coded as corresponding to inference and no-inference interpretations as described: Target choices compatible with the inference were faster than those only compatible with a no-inference reading, and Covered Picture choices based on the falsity of the inference were slower than Covered Picture choices (which could be) based on the falsity of literal meaning alone. To investigate this result statistically, we analysed the data as a $2 \times 2$ interaction design, using the same statistical analyses as detailed for Experiment Ia. The detailed results are summarized in Table 3. The $2 \times 2$ interaction was highly significant, as was the relevant simple effect comparing inference vs. no-inference responses for Target choices. For Covered Picture choices, there was a numerical effect in the same direction as for SIs (Inf $>$ NoInf), but this did not reach significance.

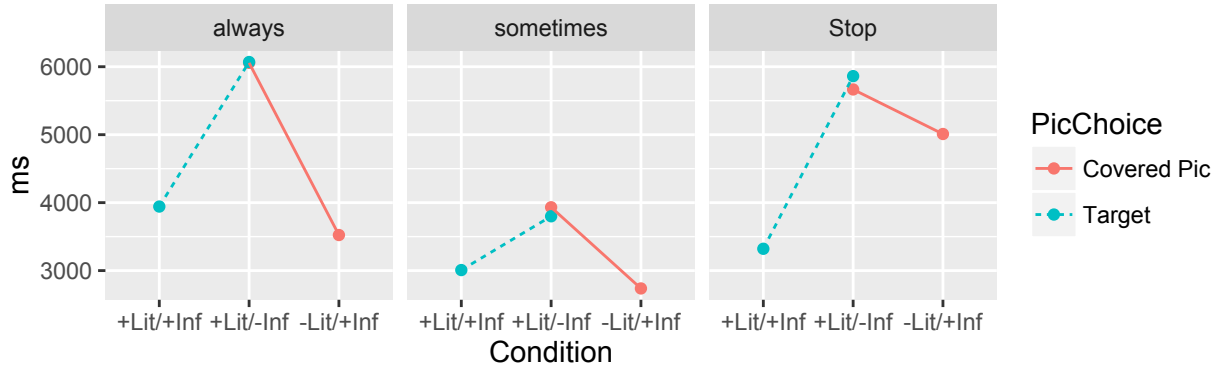



**Figure 7.** RTs for responses by picture choice and inference status for *stop* data. RTs for *always* and *sometimes* from Experiment Ia repeated for comparison. +LIT/+INF target choices and +LIT/-INF Covered picture choices are taken to reflect inference interpretations, and +LIT/-INF target choices and -LIT/+INF Covered picture choices no-inference interpretations.

| P's | $\beta$ | $SE$ | $t$ | $\chi^2$ | $p$ |
|---|---|---|---|---|---|
| **Interaction** | 3088.2 | 592.1 | 5.22 | 19.66 | **<.001** |
| **Simple Effects:** | | | | | |
| Covered Picture Choices: inference > no-inference | -772.9 | 515.5 | -1.50 | 2.16 | = .14 |
| Target Choices: inference < no-inference | -2340.0 | 431.7 | -5.42 | 21.55 | **<.001** |

**Table 3.** Summary of response time analyses for Experiment Ib: Interaction between Picture Choice and inference status and simple effects for relevant paired factor levels.

The first finding extends the findings in Romoli and Schwarz (2015) and our Experiment Ia to the domain of presuppositions, as inference interpretations seem to be faster than no-inference ones when looking at acceptance judgments. The direction of the RT effect for Covered Picture responses seems parallel to the SI-results in Bott and Noveck (2004) and Cremers and Chemla (2014), again extended to presuppositional inferences. However, as in the case with SIs, it's worth noting that the more direct comparison with these previous studies would be between Target choices based on a no-inference interpretation and Covered Picture choices based on an inference interpretation, and we find no difference here, parallel to the case of SIs. Thus, our result here differs from both the previous findings for SIs as well as those for Ps by Chemla and Bott (2013), but the results are parallel to our findings for SIs in Experiment Ia. In sum, based on the results from Experiments Ia and Ib, we find no difference in the processing patterns (measured through RTs) of Ps, DSIs or ISIs. This is consistent with the SI approach to Ps' prediction of uniformity between SIs and Ps (i.e. (6-b)). Next we turn to investigating the effect of one more variable, that of prosody, on these inferences, as a further test of their uniformity.

### 4.3 Experiment II: The effect of prosody on inference interpretations

It has been observed in the literature that prosodic focus interacts with both SIs and Ps. In particular, in the case of ISI, stress on the scalar terms trigger has been argued to be necessary for the felicity of a reading without the inference (ie. also described as 'cancellation' of the implicature; see Horn 1989; Fox and Spector 2009 and references therein).

(42) John didn't ALWAYS go to the movies.

As for presuppositions, it has also been observed that stress on the trigger changes the availability of the inference reading (see Beaver 2010; Abusch 2002; Simons et al. 2017; Abrusán 2014; Romoli 2012; Esipova 2018). In cases of negation like (43), stress on the trigger has also been associated with less inference interpretations.

(43) John didn't STOP going to the movies.

There are ongoing debates about the precise role of prosody in cases (42) and (43) and how it interacts with the mechanisms for deriving implicatures and presuppositions. All that matters for current purposes is that according to the SI approach to Ps, we expect stress to play a parallel role for SIs and (the relevant type of) Ps. That is, on this approach the derivation of (indirect) implicatures and ('projecting') presuppositions under negation proceeds in entirely parallel ways, and thus should be modulated in the same way by variations of the prosody. A traditional approach, on the other hand, can more easily accommodate a difference in the effect of prosody on the two inferences.

In order to assess this prediction, we conducted an experiment comparing written stimuli to auditory ones, which either had neutral intonation or prosodic stress placed on the expression giving rise to the implicature or presupposition. The setup is overall parallel to that above, with a sentence-picture matching task that included a Covered Picture.[18]

### 4.3.1 Methods

#### *4.3.1.1 Materials & Design*

The sentences were slight variations of those above, with a more uniform wording for the *always* and *stop*-versions:

(44) a. John didn't stop going to the movies this week.
b. John didn't always go to the movies this week.

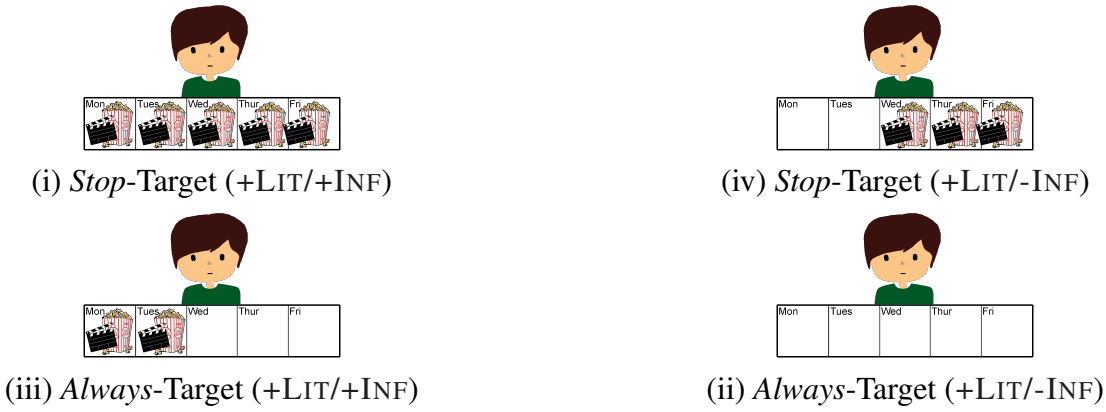
These were presented along with one of the picture variations in Figure 8 and a Covered Picture as the alternative choice. As before, the +LIT/-INF pictures can only be accepted if the judgment is based on a reading that lacks the respective inferences. In the WRITTEN condition, the sentences in (44) were presented as text on the screen. For the auditory conditions, we used audio recordings of the sentences in (44). In the NO-STRESS condition, a neutral prosody, as would be appropriate in an all-new context, was used. In the STRESS condition, *always* and *stop* bore the main pitch accent of the sentence.

In addition to 24 critical items, there were 48 fillers, 9 using *stop* with negation and Covered Picture-choices, 15 with affirmative *stop* (8 Target and 7 Covered Picture Choices), as well as 24 items replicating that pattern for *always*.

#### *4.3.1.2 Participants & Procedure*

The design was between-groups, so each participant was only exposed to one mode of presentation (WRITTEN, NO-STRESS, STRESS). The NO-STRESS data was collected as part of an eye-tracking

[18] Note that this experiment is different from the previous two in that we are no longer looking for uniformity in processing patterns. Instead we are investigating whether there is uniformity in the response of these inferences to prosodic stress, measured through rates of derivation. While the measure is different, the SI approach to Ps' prediction is similar to that made for Experiments Ia and b; namely, that there will be uniform effects of prosodic stress on the pattern of derivation rates. That is, we do not take this approach to be requiring that the effect needs to be to the same *extent* for both these inferences, just that it needs to be in the same *direction*.




(i) *Stop*-Target (+LIT/+INF)

(iv) *Stop*-Target (+LIT/-INF)

(iii) *Always*-Target (+LIT/+INF)

(ii) *Always*-Target (+LIT/-INF)

**Figure 8.** Target pictures for Experiment II.


WRITTEN NOSTRESS STRESS

**Figure 9.** Target selection rates across conditions for the WRITTEN, NO-STRESS, and STRESS variants.

experiment, but we only focus on the response patterns here.[19] A total of 97 undergraduate students from the University of Pennsylvania participated in the experiments for course credit (23 in WRITTEN, 27 in STRESS, and 47 in NOSTRESS). Instructions and practice trials were parallel to those for the previous experiments. Participants saw a total of 72 trials, and the 4 conditions of the 24 critical items were counter-balanced across groups of participants.

### 4.3.2 Results & Discussion

The dependent variable of main interest for this study was response rates, as we were interested in assessing the impact of prosody on the prevalence of inference interpretations. The overall response patterns across conditions are illustrated in Figure 9. The key observation is that we find variation in the frequency of target choices in the +LIT/-INF condition across different stimulus presentation types. In the NOSTRESS condition with auditory stimuli using neutral prosody, target acceptances seem to be lower than in the WRITTEN condition, indicating a greater prevalence of inference interpretations, for both *always* and *stop*. However, in the STRESS condition, we find the opposite effect for *stop*, as the marked prosody increased the availability of no-inference interpretations.

To assess the main contrasts of theoretical interest statistically, we conducted 2×3 mixed-effect model logistic regression analyses using treatment coding on the data for the +LIT/-INF conditions, with varying baselines to assess different simple effects. Comparing the WRITTEN version to the NOSTRESS version

[19] As will be detailed below, there were very few Target choices in the +LIT/-INF condition for *stop* here, which prevented any meaningful eye tracking data analysis for the trials of interest.

confirmed a significant decrease in Target-acceptances for both *stop* ($\beta = -4.85$, $SE = 1.23$, $z = -3.96$, $p < .001$) and *always* ($\beta = -3.98$, $SE = 1.18$, $z = -3.36$, $p < .001$). The interaction term for this comparison did not reach significance ($p = 0.12$), but there is a significant simple effect with fewer Target acceptances for *stop* than for *always* in the NoStress condition ($\beta = 1.42$, $SE = 0.40$, $z = 3.53$, $p < .001$). Turning to a comparison of the Written condition and the Stress condition, there was a significant increase in Target acceptances for *stop* ($\beta = 2.49$, $SE = 1.23$, $z = -2.03$, $p < .05$), and a marginally significant decrease for *always* ($\beta = -2.39$, $SE = 1.25$, $z = -1.91$, $p < .1$). In addition, there was a significant interaction ($\beta = -4.89$, $SE = 0.69$, $z = -7.07$, $p < .001$). Comparing the Stress and NoStress conditions directly revealed more Target acceptances for *stop* sentences in the Stress condition ($\beta = 7.35$, $SE = 1.21$, $z = 6.07$, $p < .001$), while their was no difference between these condition for *always* sentences. Finally, the interaction term for this comparison was also significant ($\beta = 5.76$, $SE = 0.70$, $z = 8.21$, $p < .001$).

The outcome pattern for the prosodic manipulations is striking, and entirely unexpected from the perspective of the SI approach to Ps, at least in the simple version we are focusing on here. If presuppositions and implicatures are derived in parallel ways based on reasoning over alternatives, then prosodic stress on the inference-triggering expression should have parallel effects. However, for *always*, we find that auditory stimuli in general increase the availability of inference interpretations. And at least numerically, in our results stress increases the likelihood of inference interpretations for implicature-triggers rather than decreasing it (although this effect did not come out as significant in our analyses).[20] The effects for *stop*, on the other hand, go in opposite directions based on whether it is stressed or unstressed in the auditory versions. The latter leads to an increase in inference interpretations, whereas the former leads to a decrease. This last result is in line with the observations in the literature mentioned above, about stress on presuppositional trigger leading to an increase in no-inference interpretations. Most important for our purposes is the different effect of prosody on SIs and Ps, which is unexpected by the SI approach to Ps.

This difference in the effect of prosody on SIs and Ps provides a first clear argument against a unified analysis of the derivation of these inferences. In contrast, these results are perfectly compatible with a more traditional view that sees them as theoretically very different cases. The next section presents further evidence along the same lines, produced as a result of evaluating the other identified prediction made by the SI approach to Ps. Namely, that in affirmative contexts, Ps and entailments should behave uniformly (i.e. (6-a)).

Before that, however, let us mention briefly how these results relate with the work on 'scalar diversity' done by Van Tiel et al. (2016) (among others). This work has shown substantial variation in the derivation rates of different scalar implicatures. One might wonder whether the difference we have found between SIs and Ps might 'just' be a sign of this scalar diversity, rather than evidence of different derivational mechanisms. However, the fact that the prosodic stress appears to have, not just *different*, but *opposite* effects on the derivation rates of these inferences is more in-line with a qualitative distinction between them (à la different derivational mechanisms), than a quantitative difference (à la scalar diversity).

[20] Note however that this result is still compatible with the claim in the literature that stress on the trigger is a necessary but not sufficient condition for the no-inference interpretation to become available.

## 4.4 Experiment IIIa: *Stop* in affirmative sentences

### 4.4.1 Motivations

We set out to test the predictions of the SI approach to Ps, as presented in (6-a) and (6-b). Turning to the former, the approach sees Ps as simple entailments. This feature of SI approaches to Ps predicts that - everything else being equal - the inference traditionally considered to be a P should be entirely on par with other entailed content (6-a). That is, they predict uniformity between Ps and simple entailments in affirmative contexts. For example, according to the SI approach to Ps, *stop* in the following sentence is assumed to entail (and only to entail) both of the following:

(45) John stopped going to the movies on Wednesday.
  a. John did not go to the movies from Wednesday on.
  b. John did go to the movies before.

Both these inferences are derived from the same sentence and, according to the SI approach to Ps, they are equivalent in status (i.e. they are both simply entailed). As a result, we take it that the SI approach to Ps would predict a greater degree of uniformity in the behavior of these inferences, compared to others we have investigated thus far. In particular, we take it that the SI approach to Ps predicts that rejecting a picture based on one of these should be just as fast as for the other. In contrast, traditional accounts posit that while both (45-a) and (45-b) are entailed by (45), (45-b) is also presupposed by (45) and thus differs in status from the first. More precisely, the fact that (45-b) is both entailed and presupposed might lead to different patterns in behavioral data than (45-a), which is simply entailed (see Kim 2007 and Schwarz 2016b for previous instances of this approach to *only* and definites, respectively). We investigated the relationship between rejections based on either one of these two inferences in affirmative sentences.

### 4.4.2 Methods

#### *4.4.2.1 Materials & Design*

The materials of this experiment were part of the same overall experiment reported as Experiment Ib on *stop* in negative sentences above. Affirmative sentences with a presupposition trigger such as *stop* differ from those with DSIs in that they cannot be judged true in a context where the inference of interest (that the relevant activity had been going on before) is false. This renders such sentences unsuitable for a direct comparison with affirmative SI sentences (i.e., DSIs), but they provide a possible angle for assessing the status of the inference. Note first that rejection responses in such contexts are captured on both traditional accounts and the SI approach to Ps, though in different ways: the former sees it as a case of presupposition failure, whereas the latter sees it as a simple rejection based on unmet entailments. The contexts we used are depicted in Fig. 10. In the -LIT/+INF condition, the overt picture does not match the sentence based on its simply entailed content, since the movie-going continued past Wednesday, but the inference that John was going to the movies before Wednesday is met. In contrast, in the +LIT/-INF condition, the inference — be it both a presupposition and an entailment, or merely an entailment — is not met, while the simply entailed content, that there was no 'movie-going' after Wednesday, does hold.

#### *4.4.2.2 Participants & Procedure*

The data stem from the same 34 participants as in Experiment Ib, and the sentence-picture combinations that they saw were variants of the negative versions reported there. In particular, subjects saw 6 sentences in the -LIT/+INF condition and 6 in the +LIT/-INF condition, drawn from a total of 24 sentences,


-LIT/+INF +LIT/-INF

John stopped going to the movies on Wednesday.

**Figure 10.** Visual stimuli for inference vs. basic entailment-based rejections of *stop* in affirmative contexts

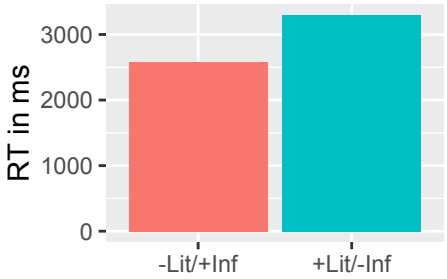



**Figure 11.** Experiment II ('stop') RTs for rejections in the Inference False and Inference True conditions.

counterbalanced across groups as described above. The Instructions and procedure were as laid out for Experiment Ib, (see section 4.2.1).

### 4.4.3 Results & Discussion

Unsurprisingly, Covered Picture selections were at ceiling level (over 97% for both conditions). RTs are illustrated in Fig. 11. Covered Picture choices were slower in the +LIT/-INF condition (3296ms) than in the -LIT/+INF condition (2583ms). This difference was statistically significant, as confirmed by a mixed-effect regression analysis with random effects for subjects and items, including intercepts and slopes ($\beta = -689.6$, $SE = 203.1$, $t = -3.40$, $\chi^2 = 9.48$, $p < .01$).

The observed difference in RTs points to a difference between the two ingredients of meaning at play. This pattern is not predicted by the SI approach to Ps, which would expect uniformity between these conditions, (6-a). On the other hand, it fits quite naturally with a traditional account, where one is presupposed and entailed, while, the other is simply entailed. Previous findings by Kim (2007) and Schwarz (2016b) have shown that rejection of sentences based on presupposed material is slower than rejection based on entailed content, and the present results fits into that picture straightforwardly on the traditional view. The SI approach to Ps does not offer an obvious explanation for this difference, as it sees both aspects of the meaning of (45) as simple entailments. However, one way of potentially saving the SI approach to Ps would be to challenge the assumption implicit in this interpretation of the data, namely that entailments of a sentence (that are generally comparable, specifically with regards to the task at hand), are on par with one another, specifically with respect to behavioral patterns such as those in RT results. An obvious approach to test this in light of our previous comparisons between *always* and *stop* is to look at different falsifying scenarios for the former. If we also find a difference between corresponding entailments associated with


**Figure 12.** Follow-up experiments on rejections of *always*-sentences

sentences containing *always*, then our current result for sentences containing *stop* would be less problematic for the SI approach to Ps.

### 4.5 Experiment IIIb and c: Rejections of *always* based on different entailments

When we compared sentences with *always* to ones with *stop* under negation, there were two ingredients of the overall conveyed meaning, which differed in status when occurring under negation:

(46) John didn't always go to the movies.
  a. There were times when John did not go to the movies.
  b. John sometimes went to the movies.

The inferences in (46-a) and (46-b) are traditionally analyzed as an entailment and an SI, respectively. However, in the case of an affirmative *always* sentence like (47) both (46-b) and the negation of (46-a) (i.e. (47-a)) are entailed. This makes affirmative sentences like (47) a good test for the assumption that different aspects of the entailments of a sentence yield equivalent RT results when providing the grounds for rejection of the sentence.

(47) John always went to the movies.
  a. It's not the case that there are times when John did not go to the movies.

Two follow-up experiments looked at rejections of positive *always*-sentences based on pictures corresponding to the two entailments in question. The design is illustrated in Fig. 12.

The crucial manipulation was whether the *always* sentence was falsified by an overt picture where the depicted individual sometimes went to the movies or whether they never went to the movies. If the two different aspects of the overall entailments of the sentences involved an asymmetry parallel to that found for the two ingredients of *stop*-sentences, then we would expect a similar RT-difference between the two conditions. In contrast, if no such difference is involved, we expect no RT-contrast, and an interaction with the results for *stop*. The latter prediction was borne out. RTs for the ALWAYS PICTURE (2383ms) and the NEVER PICTURE (2321ms) did not differ significantly from one another. Comparing the results statistically to those for *stop* reported above (analyzed as a between-subjects, within-items design with a maximal random effects structure for the latter) yielded a significant interaction ($\beta = 743.1$, $SE = 224.5$, $t = -3.31$, $\chi^2 = 9.12$, $p < .01$).

A potential concern about this first follow-up is that it involved empty calendar slots. In particular, one might worry that the NEVER PICTURE version, which conceptually corresponded to the more difficult *stop*-condition with an unmet presupposition, might lend itself to a relatively easy task-strategy of rejection based on the completely empty calendar strip, thus hiding potential delay effects. A second follow-up addressed this issue by filling the relevant calendar slots with another image type instead (see right side of Fig 12). While there was a small numerical difference between the ALWAYS PICTURE (5505ms) and the NEVER PICTURE (5735ms) in the results of this experiment, the difference was not statistically significant.[21] Comparing these results to the data obtained for *stop* from above, we again find a statistical interaction ($\beta = 156.13$, $SE = 72.93$, $t = -2.14$, $\chi^2 = 4.48$, $p < .05$)

What both of these follow-ups suggest, then, is that while there is an asymmetry in the role of the two inferences in question in the case of *stop*, this is not the case for the different aspects of the entailments of *always*. While this of course does not conclusively show that all entailments have the same processing status, it further suggests that in the case of *stop*, we are not dealing with two aspects of the overall entailment, as posited by the SI approach to Ps. In contrast, these results are consistent with the traditional perspective that the relevant inferences associated with affirmative *stop* sentences (i.e. (45)) have different statuses (i.e. simply entailed vs. entailed and presupposed).

## 5 GENERAL DISCUSSION

We set out to investigate the SI approach to Ps by trying to answer the main question outlined in (48). The predictions of the SI approach to Ps in regards to this question are repeated in (49-a) and (49-b). Experiment Ia, Ib and II set out to test prediction in (49-b). Experiments IIIa-c tested the prediction in (49-a).

(48) **Main question**: Do behavioral patterns in experimental data, e.g., in terms of (RTs) and response patterns, yield evidence for a distinction between Ps and entailments in affirmative contexts and between Ps and SIs in other contexts?

(49) **Predictions**: All else being equal,

a. In affirmative contexts, Ps and entailments should behave uniformly.
b. In all other contexts, Ps and SIs should behave uniformly.

First, we will focus on Experiments Ia and Ib, as these produced results that were consistent with the prediction in (49-b). Following this, we will consider the other experiments, which produced results that were not in line with the predictions in (49-a) and (49-b), and discuss the challenge they pose for the SI approach to Ps.

### 5.1 What doesn't challenge the SI approach to Ps

To briefly recap the situation in the literature, the classic finding since Bott and Noveck (2004) is that rejecting a sentence when its SI is false takes more time than accepting it. The same paradigm was then applied to Ps by Chemla and Bott (2013) and they found the opposite result: rejecting a negated sentence whose presupposition is not globally met takes *less* time than accepting it. On the basis of this result,

[21] Note that the overall longer RTs here are due to a slight variation in task, where a context sentence was included and the events in the calendar were revealed in two steps. Since the main measures of interest are a comparison between the two *always*-conditions and the interaction, this main effect of the task does not affect the interpretation of the results for our purposes.

Chemla and Bott (2013) concluded that Ps, unlike SIs, are not associated with a delay and that the answer to the question in (48) is positive: the processing of Ps and SIs is different, which in turn is a challenge for unified accounts like the SI approach to Ps. On the other hand, Romoli and Schwarz (2015) found that accepting negated sentences with a true presupposition is faster than accepting it when its P is not satisfied in the context, and they found parallel results for SIs, with faster acceptance of inference interpretations than no-inference interpretations. On the basis of this result, these authors concluded that there is no clear overall evidence for either SIs or Ps being associated with a delay or for the two inferences being different. On the face of it, the results from these two studies appear in conflict and they seem to give us opposite answers to the question of whether Ps and SIs differ. However, there is an obvious difference between these studies, which could account for the different results produced. Specifically, the two studies looked at different comparisons across acceptance and rejection responses; while Chemla and Bott (2013) compared acceptance versus rejection responses of the same item, Romoli and Schwarz (2015) compared acceptance versus acceptance responses across different items. Gaining a comprehensive comparative perspective required looking at both acceptance and rejection responses systematically, and this constituted the main motivation for Experiment Ia and Ib.

In Experiment Ia, we compared direct and indirect SIs using the paradigm from Romoli and Schwarz (2015), to test whether their finding was specific to indirect SIs. Moreover, we extended their approach by comparing both acceptance versus acceptance responses as well as rejection versus rejection responses across items. Both direct and indirect SIs yielded faster responses in the inference condition than in the no-inference condition when we considered acceptance responses, thus replicating Romoli and Schwarz (2015) on indirect SIs and extending their results to direct ones. On the other hand, looking at rejections yielded the opposite pattern, as rejections in the inference condition were slower than in the no-inference condition. Thus, we find uniformity between direct and indirect SIs and we also reconcile the findings of Chemla and Bott (2013) and Romoli and Schwarz (2015) to some extent.[22] In Experiment Ib, we extended the same paradigm to Ps, by looking at sentences with *stop* under negation. The RT pattern was parallel to that for SIs, with a cross-over interaction reflecting opposite patterns for acceptance and rejection responses.[23]

The uniformity in the overall shape of the RT patterns of direct SIs, indirect SIs and Ps in these experiments is in line with the prediction in (49-b) and thus provides no evidence against the SIs approach to Ps. Moreover, we found no evidence for either Ps or SIs being associated with a delay in RTs, a point that we will return to in a moment.

## 5.2 What does challenge the SI approach to Ps

In Experiment II, we investigated the effect of prosody on the availability of inference interpretations for SIs and Ps. In contrast to the results from Experiment Ia and Ib, the results of Experiment II went against the prediction in (49-b). That is, Experiment II found directly opposite effects of placing prosodic stress on the inference-triggering expressions for SIs and Ps: inference rates decreased for SIs, relative to written stimuli, but increased for Ps. These results run against the SI approach to Ps' prediction of uniformity of behavior across these inferences.

[22] Note that, while as far as RTs are concerned our results are comparable for ISIs and DSIs, the rate of implicature interpretations is significantly higher for DSIs. It's possible that this is simply due to complexities introduced by negation, but a more detailed explanation will have to be fleshed out in future work.

[23] Note that these results touch on an issue that has been investigated in detail elsewhere; namely, the effect of accepting/rejecting positive/negative sentences. In general, the work in this area seems to be consistent with our results, in that, judging sentences as true has been found to take longer than judging them as false (Wason, 1959). For a recent summary of the relevant literature see Dale and Duran (2011).

With regards to the first prediction of the SI approach to Ps' (49-a), namely that in affirmative contexts, elements of meaning that have traditionally been thought of as Ps and entailments should behave uniformly. This prediction stems from the fact that the SI approach to Ps analyses the relevant inferences as simple entailments, and was addressed by Experiments IIIa-c. Experiment IIIa tested prediction (49-a) by comparing the entailment and the presupposition of 'stop' in affirmative sentences. Specifically, it compared the behavior (measured as RTs) of participants who were rejecting a picture based on the notions that something was happening before or that it is not happening any longer, respectively. As the SI approach to Ps treats both of these elements of meaning as simple entailments, it did not predict a difference in RT behavior between these conditions. On the other hand, the traditional approach makes no specific predictions in regard to this comparison, but is perfectly compatible with there being a difference between the two. Experiment IIIa found a difference in the RTs associated with these different rejection responses, with slower responses for presupposition-based rejections, in line with previous findings (Kim, 2007; Schwarz, 2016b). This result is consistent with the traditional approach to Ps, but is a challenge for the SI approach to Ps. One way the SI approach to Ps could overcome this challenge would be to argue that not all simple entailments are on a par with one another with regard to RT behavior patterns, and so, Experiment IIIa's result should not be taken as indicative of a difference in their nature (i.e., they could still both be simple entailments of 'stop'). Experiment IIIb and IIIc set out to explore this proposal by comparing the RTs associated with rejections based on two elements of meaning that have both been traditionally analysed as simple entailments of 'always'. These experiments found no difference in the RT behavior of rejections based on these two different simple entailments. These results make the possible explanation of Experiment IIIa's results (that different simple entailments have differing RT patterns) by the SI approach to Ps less plausible. As this approach would now need to also explain why the RT behavior of the simple entailments of 'stop' differed, while those of 'always' did not.

It is worth considering these results in light of other recent experimental work which has also challenged the predictions of the SI approach to Ps. In particular, two other recent studies investigated the prediction in (49-b) by looking at how different populations interacted with these elements of meaning, using a Covered Picture selection task parallel to the one employed in the experiments reported here. Bill, Romoli, Schwarz, and Crain (2016) and Kennedy, Bill, Schwarz, Crain, Folli, and Romoli (2014) find that healthy adults, children (ranging from 4-7), and individuals with Broca's Aphasia (BAs) relate to Ps and SIs differently. Healthy adults and BAs tend to respond based on an inference reading when responding to sentences associated with SIs, while children are more likely to access an no-inference reading. In contrast, for presuppositions, children and BAs pattern together and are more likely than healthy adults to respond based on an inference interpretation. Regardless of the exact explanation for each population's behavior in the respective cases, the fact that we get a dissociation in the patterns across populations, in particular with the BAs patterning with different groups for Ps and SIs, goes against the prediction in (49-b). Therefore, these results, combined with our present results provide strong evidence against treating SIs and Ps in an entirely uniform manner.

### 5.3 Are SIs (and Ps) associated with RT delays?

Results such as those found by Bott and Noveck (2004) are commonly interpreted to indicate that implicatures require a costly computation that lead to delays in processing (Bott and Noveck, 2004; Huang and Snedeker, 2009a; Bott et al., 2012). Our results, on the other hand, did not involve a general delay in the inference conditions, for either SIs or Ps. In particular, when comparing acceptance judgments in Experiment Ia and Ib, cases where the Target picture was compatible with the inference interpretation were faster than ones where it was only compatible with the no-inference interpretation. This is incompatible

with an account that simply posits two stages — an initial stage where only the literal meaning is available, and a later stage, where the inference interpretation is available — and maps these onto response time results. Both of the visible pictures involved in the acceptance comparison are compatible with the literal meaning, and thus should yield equivalent response patterns (or, if anything, a delay in the inference condition). In contrast with the acceptance comparison, the comparison of rejection responses yielded a pattern where responses based on an inference interpretation were slower. On their own, these might be seen as compatible with an account based on processing delays for inference interpretation. But given the cross-over interaction in our results, an alternative explanation of the effects is called for.

In the following, we sketch how the RT patterns in our data can be captured in terms of a conflict between pragmatic principles. To begin with, the relatively rapid acceptances based on inference interpretations suggests that the inferences are readily available. But why should the acceptance of pictures that are only compatible with a no-inference interpretation be slower? It cannot be due to a delay in availability of the no-inference interpretation since a), the inference interpretation entails the no-inference interpretation and b) rejections of pictures based on the no-inference reading are fast. An alternative explanation of the overall pattern in our data starts from the observation that delays arise precisely in those circumstances where the inference and no-inference interpretations conflict with one another. For example, we find relatively slow Target picture acceptances when the target is compatible with the no-inference interpretation but incompatible with the inference interpretation (Fig. 3a for DSIs, Fig. 3c for ISIs, and Fig. 6b for Ps). Similarly, Covered Picture selections are also slow in the very same circumstances. One possibility then, is that there are opposing pressures favoring the respective interpretations, and that delays arise precisely when there is a conflict between these factors. More specifically, we assume that comprehenders follow a general principle of charity, i.e., they generally try to construe utterances in such a way that they are true of the circumstances at hand. In our case, charity can plausibly be seen as corresponding to selecting the Target picture, as that is the obvious and salient option at hand. On the other hand, it is intuitively plausible that inference interpretations are generally preferred. For SIs, this is in line with naive speakers' intuitions about the meaning of *some*.[24] For Ps, a preference for an inference interpretation is in line with the common claim in the literature that interpretations including presuppositions seem to be the clear default, whereas no-inference interpretations are often thought to only be marginally available.

In sum, we assume the following two principles at work:

(50) **Charity**: Construe sentences as true if possible.[25]

(51) **Inference preference**: Inference interpretations are preferred (for both SIs and Ps)

The pressures of selecting the Target picture and the preference for inference interpretations oppose one another in precisely those conditions where we find a RT delay in our data. In the +LIT/-INF conditions, the principle of charity favors the Target picture, and the preference for inference interpretations favors the Covered Picture. Whether participants end up choosing the Target or the Covered Picture, their responses are delayed in these cases, compared to Covered Picture and Target picture selections in the relevant control conditions.[26] It is interesting to relate this account to an idea presented by Katsos and Bishop (2011), who

[24] Indeed, as anyone that has taught introductory logic can confirm, it takes substantial effort to convince students that *some*-statements are in principle compatible with universal scenarios, i.e., that *some* does not literally mean *some but not all*.

[25] In our set-up, this plays out as a pressure to select the Target picture, if possible.

[26] Note that, as RT-measurements are a relatively late and global measure of linguistic processing, our results do not preclude the possibility of there also being an initial delay associated with SI derivation, as found in studies measuring online processing more directly, such as Huang and Snedeker (2009b) and others. Thanks to Jesse Snedeker for discussion on this point.

explain acquisition data in terms of pragmatic tolerance: from our perspective, one could see this in terms of the charity principle being stronger in children than the preference for inference interpretations.

## 6 CONCLUSION

Recent proposals in the theoretical literature have put forth a unified view of a variety of inferences that traditionally have been seen as falling into different classes, under the umbrella of SIs. A simple and powerful approach to investigating these unified proposals experimentally is to compare the inferences in question directly to one another, using behavioral measures. Everything else being equal, unified accounts predict uniform behavior. This approach has been applied fruitfully to the case of free choice inferences (Chemla and Bott 2014; Tieu et al. 2015b) and multiplicity inferences (Tieu et al. 2015a), among others. We applied it to the comparison between classical SIs and Ps to investigate the uniformity prediction of recent SI approaches to Ps (Chemla 2009; Romoli 2015 among others). Previous results from the literature (Chemla and Bott, 2013; Romoli and Schwarz, 2015) bearing on this issue have yielded conflicting results. We proposed that the different results were due to differences in terms of what types of responses (in terms of acceptances vs. rejection responses) were compared. Our first few experiments (Ia & Ib) show that, once the acceptance vs rejection pattern is factored in, then, in regards to the processing patterns, there is no longer any clear evidence for differences between the inference types. Furthermore, these results challenge the common interpretation of previous RT findings that implicatures are associated with an RT-delay due to the cost of computing these inferences online, and we sketched an alternative perspective based on our results. However, when we turned to Experiment II, we found that, counter to the predictions of the SI approach to Ps, there was a difference in the way these inferences were affected by prosody. In Experiment IIIa, we tested another prediction of SI approaches to Ps, namely that the relevant inferences of sentences including triggers like *stop* are simple entailments in affirmative contexts, which (again, everything else being equal) predicts uniform behavior with other simply entailed content. The results of this experiment showed that participants were slower to select the Covered Picture based on content that is traditionally thought to be entailed and presupposed compared with content traditionally thought to be simply/only entailed. These results are not consistent with the expectations of the SI approaches to Ps. In Experiments IIIb and c we investigated the plausibility of a possible explanation that SI approach to Ps could use to account for the differences in Experiment IIIa; that different simple entailments might show differing RT behavior. We investigated this possible claim by comparing the RT behavior associated with two simple entailments of 'always', and found no difference between them. These results reduce the plausibility of Experiment IIIa's results being accounted for with such an explanation. So, going back to the question of whether there is evidence from processing for a difference between SIs and Ps, we can now give it a positive answer: there is evidence for a difference between Ps and SIs. The first piece of evidence being the difference in the way Ps and SIs interact with prosody, and the second being the difference in how Ps and simple entailments are treated in affirmative sentences. Finally, our results link up quite nicely with recent evidence from the study of language acquisition (Bill, Romoli, Schwarz, and Crain, 2016) and Broca's Aphasia (Kennedy, Bill, Schwarz, Crain, Folli, and Romoli, 2014), which also produced results differentiating SIs and Ps in terms of responses patterns across populations. Considering these past findings, as well as our current results, it would appear that the SI approach to Ps is faced with a genuine challenge.

## CONFLICT OF INTEREST STATEMENT

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## AUTHOR CONTRIBUTIONS

C.B, J.R, and F.S equally contributed to designing and implementing all the reported experiments, as well as to writing this paper. C.B and J.S oversaw data collection for Experiment Ia, and F.S for Experiments Ib, II, and IIIa-c. F.S handled the statistical analyses of the data.

## FUNDING

We gratefully acknowledge support from NSF-grant BCS-1349009 to Florian Schwarz.

## ACKNOWLEDGMENTS

[To be filled in]

## SUPPLEMENTAL DATA

## DATA AVAILABILITY STATEMENT

The datasets [GENERATED/ANALYZED] for this study can be found in the [NAME OF REPOSITORY] [LINK].

## REFERENCES

Abrusán, M. (2011). Triggering verbal presuppositions. In *Semantics and Linguistic Theory (SALT) 20*, ed. D. Nan Li & Lutz (Vancouver, British Columbia), 684–701

Abrusán, M. (2014). Disappearing acts of presuppositions: Cancelling the soft-hard distinction. Unpublished manuscript CNRS

Abusch, D. (2002). Lexical alternatives as a source of pragmatic presupposition. In *Semantics and Linguistic Theory (SALT) 12*, ed. B. Jackson. 1–19

Abusch, D. (2010). Presupposition triggering from alternatives. *Journal of Semantics* 27, 1–44

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

Bates, D. M. (2005). Fitting linear mixed models in r. *R News* 5, 27–30

Beaver, D. (2001). *Presupposition and Assertion in Dynamic Semantics* (Stanford University: CSLI Publications)

Beaver, D. (2010). Have you noticed that your belly button lint colour is related to the colour of your clothing? In *Presuppositions and Discourse: Essays Offered to Hans Kamp*, eds. R. Bauerle, U. Reyle, and T. E. Zimmerman (Crispi)

Beaver, D. and Geurts, B. (2012). Presuppositions. In *Semantics: An International Handbook of Natural Language Meaning volume 3*, eds. C. Maienborn, K. von Heusinger, and P. Portner (Berlin: Mouton de Gruyter). 2432–2460

Bill, C., Romoli, J., Schwarz, F., and Crain, S. (2016). Scalar implicatures versus presuppositions: The view from acquisition. *Topoi* 35, 57–71. doi:10.1007/s11245-014-9276-1

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

Bott, L. and Noveck, I. (2004). Some utterances are underinformative. *Journal of Memory and Language* 51

1004 Breheny, R., Ferguson, H. J., and Katsos, N. (2013). Investigating the timecourse of accessing
1005 conversational implicatures during incremental sentence interpretation. *Language and Cognitive*
1006 *Processes* 28, 443–467. doi:10.1080/01690965.2011.649040
1007 Breheny, R., Katsos, N., and Williams, J. (2006). Are generalised scalar implicatures generated by default?
1008 an on-line investigation into the role of context in generating pragmatic inferences. *Cognition* 100,
1009 434–463. doi:10.1016/j.cognition.2005.07.003
1010 Chemla, E. (2009). Presuppositions of quantified sentences: Experimental data. *Natural Language*
1011 *Semantics* 17, 299–340
1012 Chemla, E. (2010). Similarity: towards a unified account of scalar implicatures, free choice permission and
1013 presupposition projection. Unpublished manuscript
1014 Chemla, E. and Bott, L. (2013). Processing presuppositions: Dynamic semantics vs pragmatic enrichment.
1015 *Language and Cognitive Processes* 38, 241–260
1016 Chemla, E. and Bott, L. (2014). Processing inferences at the semantics/pragmatics frontier: disjunctions
1017 and free choice. *Cognition* 130, 380–396
1018 Chemla, E. and Singh, R. (2014). Remarks on the experimental turn in the study of scalar implicatures
1019 (part i and ii). *Language and Linguistics Compass*
1020 Chevallier, C., Noveck, I., Nazir, T., Bott, L., Lanzetti, V., and Sperber, D. (2008). Making disjunctions
1021 exclusive. *Quarterly Journal of Experimental Psychology* 61, 1750–1761
1022 Chierchia, G. (2004). Scalar implicatures, polarity phenomena, and the syntax/pragmatics interface.
1023 In *Structures and Beyond: The Cartography of Syntactic Structures*, ed. A. Belletti (Oxford: Oxford
1024 University Press), vol. 3. 39–103
1025 Cremers, A. and Chemla, E. (2014). Direct and indiretc scalar implicatures share the same processing
1026 signature. In *Pragmatics, Semantics and the case of scalar implicatures*, ed. S. Pistoia Reda (Basingstoke:
1027 Palgrave Macmillan), Language and Cognition. 201–240
1028 Dale, R. and Duran, N. D. (2011). The cognitive dynamics of negated sentence verification. *Cognitive*
1029 *science* 35, 983–996
1030 Esipova, M. (2018). Focus on what's not at issue: Gestures, presuppositions, appositives under contrastive
1031 focus. In *Proceedings of Sinn und Bedeutung (SUB) 22* (ZAS Papers in Linguistics). 367–384
1032 Foppolo, F. and Marelli, M. (2017). No delay for some inferences. *Journal of Semantics* 34, 659–681
1033 Fox, D. and Spector, B. (2009). Economy and embedded exhaustification. Talk given at Cornell University
1034 Gamut (1991). *Logic, Language and Meaning* (University of Chicago Press)
1035 Gazdar, G. (1979). *Pragmatics: Implicature, Presupposition, and Logical Form* (New York: Academic
1036 Press)
1037 Grice, P. (1975). Logic and conversation. In The Logic of Grammar, D. Davidson and G. Harman (eds),
1038 Encino, CA: Dickenson, 64-75.
1039 Grodner, D. J., Klein, N. M., Carbary, K. M., and Tanenhaus, M. K. (2010). "Some," and possibly all,
1040 scalar inferences are not delayed: Evidence for immediate pragmatic enrichment. *Cognition* 116, 42–55.
1041 doi:10.1016/j.cognition.2010.03.014
1042 Heim, I. (1982). *The Semantics of Definite and Indefinite Noun Phrases*. Ph.D. thesis, University of
1043 Massachusetts, Amherst
1044 Heim, I. (1983). On the projection problem for presuppositions. In *Proceedings of WCCFL 2*, ed. D. P.
1045 Flickinger (Stanford University, Stanford, California: CSLI Publications), 114–125
1046 Heim, I. and Kratzer, A. (1998). *Semantics in Generative Grammar* (Malden, MA: Blackwell)
1047 Horn, L. (1972). *On the Semantic Properties of Logical Operators in English*. Ph.D. thesis, UCLA
1048 Horn, L. (1989). *A Natural History of Negation* (Chicago: University of Chicago Press)

Huang, Y., Spelke, E., and Snedeker, J. (2013). What exactly do number words mean? *Language Learning and Development* 9, 105–129

Huang, Y. T. and Snedeker, J. (2009a). Online interpretation of scalar-quantifiers: insight in the semantics-pragmatics interface. *Cognitive Psychology* 58

Huang, Y. T. and Snedeker, J. (2009b). Online interpretation of scalar quantifiers: Insight into the semantics-pragmatics interface. *Cognitive psychology* 58, 376–415

Karttunen, L. (1974). Presupposition and linguistic context. *Theoretical Linguistics* 1, 181–194

Katsos, N. and Bishop, D. V. (2011). Pragmatic tolerance: implications for the acquisition of informativeness and implicature. *Cognition* 120, 67–81

Kennedy, L., Bill, C., Schwarz, F., Crain, S., Folli, R., and Romoli, J. (2014). Scalar implicatures vs presuppositions: The view from broca's aphasia. In *Proceedings of NELS 40* (Amherst, MA: GLSA)

Kim, C. (2007). Processing presupposition: verifying sentences with 'only'. *31st Penn Linguistics Colloquium*

Magri, G. (2010). *A theory of individual-level predicates based on blind mandatory scalar implicatures*. Ph.D. thesis, Massachusetts Institute of Technology

Romoli, J. (2012). *Soft but Strong: Neg-raising, Soft Triggers, and Exhaustification*. Ph.D. thesis, Harvard University

Romoli, J. (2015). The presuppositions of soft triggers are obligatory scalar implicatures. *Journal of semantics* 32, 173–219

Romoli, J. and Sauerland, U. (2015). Presupposition and accommodation. In *Handbook of Pragmatics*, eds. S. Barron and Yueguo (Routledge)

Romoli, J. and Schwarz, F. (2015). An experimental comparison between presuppositions and indirect scalar implicatures. In *Experimental Perspectives on Presuppositions*, ed. F. Schwarz (Dordrecht: Springer), Studies in Theoretical Psycholinguistics. 215–240

Schlenker, P. (2008). Be articulate: A pragmatic theory of presupposition projection. *Theoretical Linguistics* 34, 157–212

Schwarz, F. (2015). Introduction: Aspects of meaning in context - theoretical issues and experimental perspectives. In *Experimental Perspectives on Presuppositions*, ed. F. Schwarz (Springer). 1–38

Schwarz, F. (2016a). Experimental work in presupposition and presupposition projection. *Annual Review of Linguistics* 2, 273–292. doi:10.1146/annurev-linguistics-011415-040809

Schwarz, F. (2016b). False but slow: Evaluating statements with non-referring definites. *Journal of Semantics* 33, 177–214. doi:10.1093/jos/ffu019

Schwarz, F., Romoli, J., and Bill, C. (2015). Scalar implicatures processing: slowly accepting the truth (literally). In *Proceedings of Sinn und Bedeutung*, ed. E. C. . H. Zeijlstra (Göttingen, Germany), 553–570

Simons, M. (2001). On the conversational basis of some presuppositions. In *Semantics and Linguistic Theory (SALT) 11*, eds. R. Hastings, B. Jackson, and Z. Zvolenszky. 431–448

Simons, M., Robets, C., Beaver, D., and Tonhauser, J. (2017). The best question: Explaining the projection behavior of factives. *Discourse processes* 54, 187–206

Stalnaker, R. (1974). Pragmatic presuppositions. In *Semantics and Philosophy*, eds. M. Munitz and D. Unger (New York University Press). 197–213

Sudo, Y. (2012). *On the semantics of Phi features on pronouns*. Ph.D. thesis, MIT

Tieu, L., Romoli, J., Peng, Z., and Crain, S. (2015a). Children's knowledge of free choice inferences and scalar implicatures. *Journal of Semantics* doi:doi:10.1093/jos/ffv001

Tieu, L., Romoli, J., Zou, P., and Crain, S. (2015b). Children's knowledge of free choice inferences and scalar implicatures. *Journal of Semantics*

van der Sandt, R. (1992). Presupposition projection as anaphora resolution. *Journal of Semantics* 9, 333–377

Van Tiel, B., Van Miltenburg, E., Zevakhina, N., and Geurts, B. (2016). Scalar diversity. *Journal of Semantics* 33, 137–175. doi:10.1093/jos/ffu017

von Fintel, K. (2008). What is accommodation, again? *Philosophical Perspectives, 22, Philosophy of Language*

Wason, P. C. (1959). The processing of positive and negative information. *Quarterly Journal of Experimental Psychology* 11, 92–107

Wilson, D. (1975). Presupposition, assertion, and lexical items. *Linguistic Inquiry* 6, 95–114