Investigating the timecourse of accessing conversational implicatures during incremental sentence interpretation

Richard Breheny [1]

Heather J Ferguson [2]

Napoleon Katsos [3]

[1] University College London, UK

[2] University of Kent, UK

[3] University of Cambridge, UK

Correspondence to:
Richard Breheny email: r.breheny@ucl.ac.uk
Tel: +44 (0) 20 7679 4039

UCL Division of Psychology and Language Sciences
Chandler House
2 Wakefield Street
London WC1N 1PF
UK

Abstract

Many contextual inferences in utterance interpretation are explained as following from the nature of conversation and the assumption that participants are rational. Recent psycholinguistic research has focussed on certain of these 'Gricean' inferences and have revealed that comprehenders can access them in online interpretation. However there have been mixed results as to the time-course of access. Some results show that Gricean inferences can be accessed very rapidly, as rapidly as any other contextually specified information (Sedivy, 2003; Grodner, Klein, Carbery, & Tanenhaus, 2010); while other studies looking at the same kind of inference suggest that access to Gricean inferences are delayed relative to other aspects of semantic interpretation (Huang & Snedeker, 2009; in press). While previous timecourse research has focussed on Gricean inferences that support the online assignment of reference to definite expressions, the study reported here examines the timecourse of access to scalar implicatures, which enrich the meaning of an utterance beyond the semantic interpretation. Even if access to Gricean inference in support of reference assignment may be rapid, it is still unknown whether genuinely enriching scalar implicatures are delayed. Our results indicate that scalar implicatures are accessed as rapidly as other contextual inferences. The implications of our results are discussed in reference to the architecture of language comprehension.

Introduction

Theories of Conversational Implicature and Cognitive Architecture

In the language processing literature, there has been much recent interest in the effects of pragmatic inference in online comprehension. That is, research into the online effects of constraints or principles of usage that determine comprehension in context. One important area of pragmatic research involves principles that are assumed to govern speaker-hearer interactions. Grice (1975) proposed that conversational interactions are constrained by maxims concerning relevance, informativeness and manner and that these maxims stem from the nature of conversation and the common presumption that parties to conversation are rational. Subsequent theories have taken different perspectives on the nature of conversation and these lead to different proposals about what principles apply (see for example, Horn, 1984; Sperber & Wilson, 1986; Levinson, 2000). However all such 'Gricean' theories attempt to derive certain patterns in how we interpret utterances by considering language use as an interaction among rational agents.

Sedivy, Tanenhaus, Chambers and Carlson (1999) and Sedivy (2003) represent important early work exploring the role of these kinds of pragmatic inference in online comprehension. Sedivy's research suggests that online interpretation can be affected by the results of 'Gricean' reasoning. In a separate line of inquiry, much attention has been focussed on whether certain frequent Gricean inferences should be seen as default, in the sense of not requiring contextual support (Levinson, 2000). In a series of response-based studies, Noveck and colleagues (Noveck & Poseda, 2003; Bott & Noveck, 2004) and Breheny, Katsos and Williams (2006) have demonstrated that even common Gricean inferences are not made by default but require contextual support. This response-based work is supported by Grodner and Sedivy (in press), which uses an incremental visual world paradigm.

Taken together, experimental research on Gricean pragmatic inference suggests that it is contextually specific and that the results of such inference can be accessed in incremental interpretation. These results then lead to a further question about the architecture and

mechanisms that lie behind access to Gricean inferences. Looking at Grice's own proposals, we see that he was concerned with inferences that enrich comprehension beyond 'What is said'; that is, beyond the semantic interpretation of an utterance in context. These enrichments are called 'Conversational Implicatures'. To illustrate this phenomenon, consider a case where you ask a colleague if he enjoyed a conference dinner and the colleague replies, 'I enjoyed the dessert.' Grice's proposals begin from the observation that the speaker communicates that they did not enjoy the rest of the meal without explicitly saying so. Thus the speaker says one thing (they enjoyed the dessert) and implies (or 'implicates') something further. Grice's theory is set up to explain such implicit aspects of meaning across a wide range of frequently occurring cases. One particularly well-studied example of conversational implicature is of the so-called scalar implicatures (Horn, 1972; Gazdar, 1979). Scalar implicature can be illustrated by considering the example in (1).

1. The man poured some of the water into the bowl.

This could easily be understood to mean that the man did not pour all of the water into the bowl. However the 'not all' implication is regarded by many as a pragmatic implicature, derivable in the same way as in the 'dessert' example mentioned above (see Geurts, 2010). The idea with (1) is that the conventional meaning of *some of the water* is such that 'what the speaker says' could be true even if the man poured all of the water into the bowl. According to most Gricean pragmatic theories, one infers that the man did not pour all of the water into the bowl by appeal to an informativeness maxim: if the speaker had known that the more informative proposition that the man poured all of the water into the bowl were true they would have said so. Not saying so means that we can rule out the stronger proposition, assuming that the speaker is a competent source on the question of what was poured where. Assuming that it is common ground that the

hearer could make this kind of inference then the implication can be taken to have been communicated.

If one considers Grice's theory in architectural terms, it's important to note that Grice (1975) proposes that one first establishes 'What is said' in context and then conversational implicatures can be derived from that. Thus Grice's theory implies that the representation of 'What is said' is not derived using his principles of conversation but implicatures are. The theory then is compatible with an architecture according to which first the linguistic form is processed and assigned an interpretation (e.g. that your colleague enjoyed the dessert) and then Gricean inferences further enrich that interpretation (that they did not enjoy the rest of the meal). Alternatively, Grice's theory could be implemented in an interactive model where processes that establish a semantic interpretation and implicatures operate in parallel and the latter are capable of influencing the former.

In contrast to Grice's own theory, other 'Gricean' theories, such as proposed by Sperber and Wilson (1986) or Neale (1992), hold that both the semantic interpretation ('What is said') and conversational implicatures are to be explained via pragmatic principles that constrain speaker/hearer interactions. So, to take the scalar implicature case, (1) for example, Sperber and Wilson (1995) suggest that their Relevance principle applies equally in determining the domain of quantification for the quantifier *some* (i.e. which water is being quantified over), and in the determination of the implicature that the man did not pour all of the water into the bowl. As the latter kind of approach does not imply any difference in the representation of semantic content ('what is said') and implicature, it is less compatible with an architecture according to which first 'What is said' is accessed and then implicatures.

The question of the timecourse of access to Gricean inference is explored in Huang and Snedeker (2009; in press) and Grodner, Klein, Carbary and Tanenhaus (2010). In these studies, participants hear an instruction like, 'Point to the girl that has all of the soccer balls' or 'Point to the girl that has some of the socks'. In the visual context for these items the referential

expression, 'the girl that has *all of the soccer balls*/ *some of the socks*' is initially ambiguous as there are two (or more) girls. In the visual context for these referential expressions, the intuition is that participants should begin to predict the referent from the onset of the respective quantifiers 'all' or 'some'. This is so since one girl is in possession of all of the soccer balls and no other girl is in possession of all of a set of objects; a second girl is in possession of half of the set of socks while no other girl is in possession of a part of a set of anything. While it seems clear that one should be able to anticipate the correct referent from the offset of either quantifier, it is widely assumed that this anticipation is based on an extra 'Gricean' pragmatic inference in the case of the 'some' expression compared to the 'all' expression. To see this, note that for both the 'some' and the 'all' descriptions, certain inferences are required to determine the respective domains of quantification since the arguments of the quantifiers ('the soccer balls/socks') are themselves definite. But in addition to this routine contextual inference, in the 'some' items, knowing the domain of quantification does not allow one to uniquely determine the referent in advance of the critical noun ('soccer/socks'). This is so since, strictly speaking, the girl that has all of the soccer balls also has some of the soccer balls. If one could make the inference that the speaker would not use the quantifier 'some' if they were referring to the girl with all of the soccer balls, then one could infer from the off-set of 'some' that the speaker is referring to the girl with some of the socks. As will be discussed shortly, the exact basis of this inference is yet to be agreed upon but it seems to be clearly 'Gricean' in the broad sense. Thus, if participants are able to predict the referent of, 'the girl with some of the socks' from after the offset of the quantifier, 'some', and before the onset of the noun, 'socks', then it is agreed they will have done so in virtue of accessing an extra contextual inference. That is, an extra inference over and above the contextual inferences required for resolving the domain of quantification in both the 'some' and 'all' cases.

Huang and Snedeker's studies confirm that the relevant inferences can be accessed in on-line comprehension, but their results suggest the effect is delayed relative to the baseline 'all'

case. In these studies the delay is considerable, approximately 1 second. These results then seem to contrast with those found in the work of Sedivy and colleagues (Sedivy et al., 1999; Sedivy, 2003) where the effects of a similar Gricean inference are seen almost immediately. Grodner et al. (2010) argue that some features of the studies in Huang and Snedeker (2009) that are not theory critical may have led to the delay in seeing the effects of the inference. Their own study using similar items to Huang and Snedeker controlled for these factors and found no delay. We shall return to discuss these divergent results below. Here we wish to highlight the nature of the pragmatic inferences involved in these studies and their relevance to the architectural issues mentioned above.

The stated intention behind the studies in Huang and Snedeker (2009, in press) is to explore the timecourse of access to scalar implicatures in order to see whether the semantic interpretation (or 'What is said') is consulted before implicatures are accessed. However, as has subsequently been recognised in Grodner et al. (2010), the pragmatic inference involved in these studies is not a case of scalar implicature. One cannot apply Grice's theory of quantity implicature or indeed, neo-Gricean theories of scalars (such as in Horn, 1989; Gazdar, 1979) to these items and yield the relevant inference.[1] Indeed it is an open question whether any conversational implicature (in Grice's sense) is involved here at all.

Grodner and colleagues argue that the effect in these studies can be explained as a case of 'Maximise Presupposition' (Heim, 1991; see Grodner et al., 2010 for details) and so, according to this account, participants are able to anticipate the correct referent from the offset of 'some' on the basis of making an inference about what is presupposed (rather than what is implicated). This is a plausible proposal. However, if it is on the right track, we could question the soundness

[1] According to Gazdar (1979), the scalar implicature for 'some of the socks' cannot be derived when that noun phrase is contained in a non-positive linguistic context. Here, the linguistic context is the definite description, 'the girl that has some of the socks' and it is non-positive. Other proposals are more liberal. For example Horn (1989) requires only a non-negative environment. In that case a scalar implicature might be derivable in this example but would be something like, 'You do not have to point to the girl with all of the socks'. As this implication is irrelevant to the task (there are no girls with all of the socks) we can conclude that it is not access to scalar implicatures that explains the effect.

of using this kind of item to explore the architectural issues mentioned above. Since the presuppositions of an utterance are generally thought to be derived or 'checked' before the semantic interpretation (what is said) is established (Gazdar, 1979; Heim, 1983) it would be plausible to assume that the architecture of language comprehension treats the pragmatic inferences that bear on presuppositions in a different way to implicatures. At least from a 'modular' perspective, one cannot compute what is implicated until one has computed what is said and, by the most widely accepted accounts of presupposition (see Beaver & Geurts, 2011 for a review), one computes presuppositions prior to, or along with, what is said. With regard to the current items, this perspective has a certain prima facie plausibility. It seems clear that one makes the 'Gricean' inference in order to speed up or facilitate the assignment of reference to the definite description. The referent of the description does not change because of the pragmatic inference; nor does the overall interpretation of the items. 'What is said' remains the same and there are no enriching conversational implicatures. At most, only the presuppositions change.

Although there may be alternatives to the 'Maximise Presupposition' account of the inferences in Huang and Snedeker's and Grodner and colleagues' studies (and these will be mentioned in the General Discussion), all such approaches explain the effect in terms of principles that affect the determination of the referents in context. It is quite plausible that, in architectural terms, pragmatic inferences that bear on reference, and hence the semantic interpretation of an utterance (or 'What is said') are built into a process that determines an initial semantic interpretation, while separate processes (perhaps based on Grice's maxims) derive implicatures. Indeed there have been independent proposals concerning pragmatic principles that bear exclusively on reference or presupposition but that are 'Gricean' in spirit, in that they rely on some kind of expectations of a rational speaker. These include Grodzinsky and Reinhart's Principle I (Grodzinsky & Reinhart, 1993) and Altmann and Steedman's Principle of Parsimony (Altmann & Steedman 1988).

The separation between processes delivering a semantic interpretation and those that determines conversational implicatures is, in outline, the conclusion that is suggested in Huang and Snedeker (2009) and challenged in Grodner et al (2010). We are arguing that the previous literature has yet to properly test this hypothesis and it is an important aim of the current study to do just that.

The Current Study

In the current study we aim to broaden research into the timecourse of Gricean inference to include genuinely additive conversational implicatures, the so-called 'scalar implicatures'. Another way in which our study differs from previous work on Gricean inference lies in our methods. All previous visual world research in this area has employed a method pioneered in Tanenhaus, Spivey-Knowlton, Eberhard and Sedivy (1995), where participants follow verbal instructions in relation to elements of the visual context. Here, instead, we adapt the 'look and listen' task developed in the work of Altmann and colleagues (see Altmann & Kamide, 1999; 2007) where eye movements around a visual scene are tracked while participants listen to some related discourse. Using this paradigm, it has been shown that eye movements reveal participants' anticipation of upcoming elements of the discourse before those elements are uttered. For example, in a scene containing a picture of an agent (e.g. a boy), an edible item (e.g. a cake) and distracters, it is found that from the offset of the verb in, 'The boy will eat the cake' participants are already anticipating reference to the cake in the scene. In follow-up work, Altmann and Kamide (2007) found that anticipation can be directed by the compositional interpretation of the previous parts of the sentence (rather than mere association between concepts associated with lexical items like 'eat' and 'cake'). In this later study, the image showed an agent (a man) in a visual context containing an empty drinking vessel (e.g. a wine glass) and a full drinking vessel (e.g. a beer glass) among other distracters. Participants either heard, 'The man will drink all of the beer' or, 'The man has drunk all of the wine'. In the future

tense condition, participants' fixations predict toward the full glass, while in the perfect tense condition, participants predict toward the empty glass, suggesting that the composition of the meaning of the auxiliaries and the verb is guiding prediction.

The look-and-listen paradigm is arguably a more neutral or conservative test of the timecourse of access to pragmatic information as the participants' only task is to listen for comprehension. By comparison, in the act-out paradigm used in previous research, participants have to first comprehend the instruction and then follow it. Thus, in the look-and-listen paradigm there is a reduced likelihood of task-related strategies that are not related to comprehension affecting anticipation by participants.

In previous work, we have demonstrated that access to quantity implicatures, as in the 'dessert' example, can guide predictive looks around a visual scene (Breheny. Ferguson, & Katsos, 2010; under review). In the current studies, we extend this paradigm to compare the time course of access to scalar implicatures derived from the use of 'some' (in (3) and (4) below) in comparison with access to contextual inferences relating the domains of quantification of both the quantifiers 'all' and 'some' in (3) and (4).

In all of our items, participants watched a short video depicting an agent transferring quantities of different items to one of two locations (trays labelled 'A' and 'B'). Upon completion of the video, participants viewed a still image of the last frame and heard an auditory description of the events that have just taken place while eye movements around the visual display were recorded. Descriptions were always fully informative of the events in the video and, as with the video, they focused on the agent transferring different things to the different locations. For example, one of our items involves a man pouring the contents of each of two jugs of water, one with slices of lime and the other with slices of orange into, respectively, a bowl on a tray labelled 'A' and a bowl on a tray labelled 'B'. In one version of this video, he pours half of the contents of each jug to the respective locations. Participants then hear the following:

2. The man has poured some of the water with limes into the bowl on tray A and some of the water with oranges into the bowl on tray B.

In this baseline condition, participants are only able to predict the target location to be mentioned in the first clause (tray A) from the offset of 'limes' since the quantifier 'some' provides no distinguishing information. However, in another version of this video item, the man pours all of the contents of the jug of water with oranges but only half the contents of the other jug. Two fully informative descriptions of this video are given in (3) and (4) below (which differ only in the ordering of the sentences):

3. The man has poured all of the water with oranges into the bowl on tray B and some of the water with limes into the bowl on tray A.

4. The man has poured some of the water with limes into the bowl on tray A and all of the water with oranges into the bowl on tray B.

In the case of (3), participants would be able to predict the target location to be mentioned in the first clause (tray B) earlier than in the baseline condition, from the offset of 'all'. When the spoken description is presented as in (4), participants hear the quantifier 'some' in the first clause. They could then reason that the speaker would not be describing the event of pouring all of the water with oranges since saying 'some' is not completely informative of that event and so the speaker must be describing the pouring of only some of the water with limes. Thus they should be able to predict the target location for the first clause from the offset of 'some' once they are able to access the implicature that he did not pour all of the contents of the relevant jug into the location about to be mentioned. By comparing anticipatory responses to descriptions as in (3) and (4) for the same set of events, our study will enable us to compare the

time course of access to inferences based on the meaning of the quantifier 'some' plus the implicature with that of inferences based on the meaning of the quantifier 'all' alone.

Method

Participants

Thirty participants from the University of Cambridge were paid to take part in the study. All were native English speakers with normal or corrected to normal vision and had no prior exposure to the experimental items.

Stimuli and Design

Twenty-four sets of experimental videos and pictures were paired with an auditory passage in one of three conditions. Table 1 and Figure 1 provide an example of such experimental sentences and the associated visual scenes[2]. Video clips were recorded in a single session involving one male and one female 'actor' and edited using Adobe Premier. Subsequent pictures were created from the final frame in the video clips. All visual images were presented on a 17 inch colour monitor in 1024 x 768 pixels resolution.

Two different video scenarios were created to provide contexts for the subsequent still visual image and auditory sentence. Both video scenarios began with two referents in the centre (e.g. a jug of water with limes in and a jug of water with oranges in) and two possible target locations (two trays, labelled A and B). We filmed the actor transfer around half of each referent to a different one of the target locations, pause and then continue to complete the transfer of the remaining quantity of one of the referents. To set up different visual contexts, participants could either be presented with the whole video, or a truncated version of the video, which stopped at the pause point (i.e. when both target locations contained half of each referent). Subsequent still

[2] Contact the authors for example video recordings.

visual scenes depicted the final state from each of these scenarios, and were created by extracting the final frame from each video clip. To prevent any systematic viewing strategies, the spatial arrangements and order of events was counterbalanced for the 'some' and 'all' picture elements across items. Thus, the order of auditory descriptions were counterbalanced so that the first occurring event was the first described event half of the time. Sound files consisted of a single sentence that provided a fully informative description of the transfer events depicted in the video, and thus set up three conditions as in Table 1: an *All* condition (i.e. participants watched the full video and descriptions referred first to the 'all' referent), a *Some early* condition (i.e. participants watched the full video and descriptions referred to the 'some' referent first), and a *Some late* condition (i.e. participants watched the truncated video, thus descriptions were referentially ambiguous). This resulted in a 1-factor within-subjects design, with three variables. The position of the critical word always occurred roughly mid-sentence.

One version of each item was assigned to one of three presentation lists, with each list containing twenty-four experimental items, eight in each of the three conditions. This ensured that each experimental item appeared once in each list, but in a different condition in each of these three lists. In addition, twenty-six filler items were added to each list. All filler videos involved a transfer action, and the items consisted of correctly matched picture-sentence pairings and varied in their use of quantifiers to describe the transfer event (i.e. they either did not include quantifiers (n = 11) or used different quantity words, e.g. 'both', 'the pair', 'a few' (n = 15)). Eight filler items used conjoined clause sentences, while the remaining eighteen were single-clause descriptions. Importantly, all fillers used the same referential targets (i.e. locations A and B) and as such were not deemed obviously distinguishable from the experimental stimuli[3]. These filler items were interspersed randomly among the twenty-four experimental trials to create a

[3] Please contact Heather Ferguson (h.ferguson@kent.ac.uk) for filler items.

single random order, which ensured even distribution of the three conditions throughout the task. At least one filler trial intervened between any two experimental trials.

---------- Table 1 about here ----------

----------Figure 1 about here ----------

Sentences were recorded independently for each item/condition combination in a single session from a female native British English speaker who was instructed to describe the events naturally, with a 'what happened' intonation. Subsequent analysis of critical items using the WASP speech filing system (see http://www.phon.ucl.ac.uk/resource/sfs/wasp.htm) revealed a focal pitch accent on the label letter ('A'/'B') in the first conjunct and contrastive accent on the quantifier in the second conjunct. Thus throughout the region of interest, which is prior to the conjunction, there are no focal pitch accents. Contact the authors for example audio recordings. The auditory files were presented as 44.1 KHz stereo sound clips via headphones connected to the eye-tracker PC. The temporal onsets of critical words were hand-coded with millisecond resolution using the GoldWave sound-editing package.

Procedure

Participants were seated in front of a 17 inch colour monitor with independent eye tracking system (Tobii X120) running at 120 Hz sampling rate. Viewing was binocular and eye movements were recorded from both eyes simultaneously. At the beginning of the experiment, and once every ten trials thereafter, the eye-tracker was calibrated against nine fixation points. This procedure took about half a minute and an entire session lasted for about half an hour.

The experiment was controlled using e-Prime (Schneider, Eschmann, & Zuccolotto, 2002). As illustrated in Figure 2, each trial began with the presentation of a single centrally-

located cross and participants were asked to fixate it for 1500msec before the trial was automatically initiated. At this point, the cross was replaced by the video depicting a transfer scenario, as described above. Video clips lasted on average 25 seconds (range = 19s to 34s) and were followed by a pause/ blank screen for 500ms. Next, the corresponding picture was presented with the target sentence (*All, Some early,* or *Some late*). The onset of the picture preceded the onset of the corresponding spoken sentence by 500ms. The picture stayed on the screen for fifteen seconds, and the corresponding sentence typically ended 1-2 seconds before the end of the trial.

----------Figure 2 about here ----------

Comprehension questions (see Appendix), followed half of the experimental and half of the filler trials. Participants did not receive feedback for their responses to these questions. Participants all scored at or above 80% accuracy on the comprehension questions.

Results and Discussion

Data Processing

Eye-movements that were initiated during the target sentence were processed according to the relevant picture and sound onsets for the purpose of aggregating the location and duration of each 120Hz (i.e. 8ms) sample from the eye tracker. For analysis, any sample that was deemed 'invalid' (e.g. due to blinks or head movements) was removed from the data. Using Matlab software, the spatial coordinates of the eye movement samples (in pixels) were then mapped onto the appropriate object regions using colour-coded bitmap templates; if a fixation was located within 20 pixels around an object's perimeter, it was coded as belonging to that object, otherwise, it was coded as background. All consecutive samples within one object region before the eyes moved to a different region were pooled into a single *gaze*. Finally, temporal onsets and

offsets of the gazes were recalculated relative to the corresponding picture onset by subtracting the picture onset from the relative gaze onsets and offsets.

To plot the difference in gazes to the critical some- and all-referents (i.e. trays A and B and their respective containers (e.g. jugs)) as a function of time, we calculated the number of fixations on the all-referent as a proportion of total fixations on the all- and some-referents (i.e. the 'all advantage score'). This was done by subtracting the probability of making a fixation on the some-referent from the probability of making a fixation on the all-referent. This results in an output that is symmetrical around zero such that a positive score reflects higher proportions of gazes on the all-referent and a negative score reflects higher proportions of gazes on the some-referent. The resulting plot (seen in Figure 3 below) allows us to examine when visual interpretations become significantly biased to either referent.

For statistical analysis, which is sensitive to the fact that the data is non-independent and not normally distributed, we used a log-ratio transformation to compare the probabilities of gazes to the critical some- and all-referents over time (Arai, van Gompel, & Scheepers, 2007). Thus, biases to each target referent was calculated using the following log-ratio measure: log(All/Some) = $ln(P_{(All)} / P_{(Some)})$, where $P_{(All)}$ refers to the probability of gazes on the all-referent (e.g. tray B containing the water with oranges) and $P_{(Some)}$ to the probability of gazes on the some-referent (e.g. tray A containing the water with limes); *ln* refers to the natural logarithm.

Then the data were synchronized on a by-trial basis relative to the onsets and offsets of individual words (see Altmann & Kamide, 2009) and analysed for every 20 ms time-slot over a time period ranging from the verb preceding the quantifier (e.g. 'pour') until 400ms *after* the offset of the disambiguating object (the 'wrap-up' region). These data for each condition are plotted in Figure 3. Note that for statistical analyses, these word regions were offset by 200ms to allow for the time it takes to program and launch an eye-movement (see Hallett, 1986). The average duration of the quantifiers, as well as mean durations for 'of the', 'water', 'with' and the target object, are detailed in Table 2 for each condition and are indicated by the dashed vertical

lines in Figure 3. The duration of the quantifiers differed significantly ($F(2, 69) = 67.78$, $p < 0.001$), with 'all' being significantly shorter than 'some' (213ms *vs.* 278ms (early) and 293ms (late) respectively). Importantly however, none of the following word lengths differed significantly across the three conditions ([verb]: $F(2, 69) = 0.24$, $p = 0.76$; 'of the': $F(2, 69) = 0.13$, $p = 0.86$; 'water': $F(2, 69) = 0.1$, $p = 0.85$; 'with': $F(2, 69) = 0.04$, $p = 0.97$; 'limes': $F(2, 69) = 1.33$, $p = 0.27$).

----------Figure 3 about here ----------

---------- Table 2 about here ----------

Main Analyses

Statistical analyses were performed on the log(all/some) scores with a 1-factor ANOVA for each analysis region, with condition (*All vs. Some early vs. Some late*) as the repeated measures factor. Table 3 displays the statistical details of the effects, allowing generalization to participants (F_1) and items (F_2), for each time window of interest. Strength of association is reported in terms of partial eta-squared ($_p\eta^2$).

---------- Table 3 about here ----------

Post-hoc analyses were then carried out to examine the strength of these biases to the appropriate referents for each condition. In other words, we used one-sample *t*-tests to determine at what point the visual preferences in each condition became significantly different from the zero 'no

bias' line[4]. The statistical results from these post-hoc tests are detailed in Table 4, below, allowing generalization to participants (t_1) and items (t_2), for each time window of interest.

---------- Table 4 about here ----------

Analyses revealed a marginally significant difference emerging between conditions during the verb (i.e. prior to quantifier onset). This difference became fully significant immediately from the onset of the quantifier and remained significant in all the subsequent regions of analysis ("in the", "water", "with", "limes", and [wrap-up]. Looking at the data, it appears that the source of these differences lies in the different visual biases elicited for *All* versus *Some early* conditions. Specifically, the quantifier "all" led participants to anticipate the 'all' referent long before the disambiguating information became available. Similarly, hearing the quantifier "some" (in the *Some early* condition) prompted participants to rapidly direct their visual attention to the appropriate target, thus anticipating the 'some' referent.

Results from the post-hoc *t*-tests indicated that even before the onset of the quantifier, participants in the *All* (marginal by participants) and *Some early* conditions held a preference to fixate the all-referent, while there was no bias in the *Some late* condition. We suggest that this early preference reflects one or more non-relevant factors, such as a preference to look at the location where more water was poured following the verb 'poured', or a preference to look at the location of the last pouring event following the verb (see Altmann & Kamide, 2009). Additionally there may be a low-level bias to attend to larger quantities (contained in the 'all' target). We will explore ways of controlling for these factors below.

During the quantifier word region, "all" led participants to maintain this clear visual bias towards the 'all' referent (compared to zero), while the equivalent bias was no longer significant following "some" in the *Some early* condition. As before, participants showed no bias to either

[4] Note that an all/some preference score of zero reflects equal proportions of looks between the two referents.

referent in the *Some late* condition in the quantifier region. Thus, following the quantifier "all", listeners correctly anticipated the event being referenced seemingly assisted by the bias that was present even before the onset of the quantity information. In contrast, there is a delay in the formation of a significant bias to the some-referent during the acoustic life of the quantifier "some". As such, it is possible that the immediate prediction of 'all' referents reflects a genuine advantage for predictions based on inferences that do not involve conversational implicature. However, it is also possible that this advantage is simply a reflection of the factors mentioned above, including low-level visual features of the display that influenced initial eye movements around the scene independently from the linguistic input. We will address these possibilities in our baseline-corrected analyses, below.

During "of the", "water", "with", and "limes" word regions, post-hoc tests revealed a continuing visual bias to the 'all' referent following "all". More interesting is the emergence of a fully consistent bias to the 'some' referent following "some" (*Some early*). Thus, this study provides evidence that comprehenders can make the 'some but not all' inference at least from the offset of hearing the quantifier 'some', and that this scalar implicature can be made online and in anticipation of further disambiguating information provided by a speaker. Once again, in these three word regions, the *Some late* condition did not elicit any bias to either visual target.

Finally, during the wrap-up region, all three conditions showed appropriate visual biases according to the available disambiguating linguistic input (e.g. "limes"/ "oranges").

To ensure that these effects were not influenced by participants learning a strategy to interpret "all" and "some" over the course of the experiment, we also ran statistical analyses to compare participants' performance in the three conditions as a function of trial order (i.e. those that were presented in the first *vs*. second half of the experiment). These analyses showed no 2-way interaction between Condition*Item order at any word-region [All Fs < 1.09]. Given these findings, we can conclude that participants did not modify their behaviour in this task as the experiment progressed and thus do not appear to have devised a strategy to facilitate prediction

based on an awareness of the quantity options available to them.

Baseline-corrected analyses

Inspection of the early visual biases to the *all* target during the verb region it is likely that some low-level features of the visual display or properties of the language input prior to the quantifier onset drew participants' attention to the 'all' referent independently of the quantifier's meaning. Indeed, Huang and Snedeker (2009) and Grodner et al. (2010) have also noted a general tendency to fixate the *all* targets over the *some* targets prior to linguistic input with a similar kind of visual display. Possible explanations for such biases have implicated low-level processing strategies, including an increased quantity of objects in 'all' compared to 'some' targets or a visual bias to the set of objects that has not been split.

In order that we get a more reliable indicator of the effect of the linguistic input, we conducted additional analyses that examined shifts of fixations over time. Specifically, we adopted a similar analysis procedure to that described by Huang and Snedeker (2009). In this procedure, trials were divided based on what the participant was fixating during the quantifier. Then we analysed trials where participants were fixating the incorrect referent during the quantifier, and calculated the probability that they switched their gaze to the correct referent (OnTarget trails) from the quantifier offset onwards. Separate analyses examined trials where the participant was already fixating the correct referent during the quantifier, and calculated the probability that they switched their gaze to the incorrect referent (OffTarget trials) from its auditory offset onwards. So, for example, if the delayed anticipation following "some" (relative to "all") is solely driven by an early low-level tendency to look at the larger quantity referent, then we would expect to find visual biases to the correct quantity referent emerging at the same rate for both *All* and *Some early* conditions in this new analysis. However, if the initial pattern of fixations reported above truly reflects a delay in accessing the scalar implicature, then we would

again expect slower latency to correct referents following the quantifier "some" compared to "all".

For each condition, we calculated the average probability of switching from an incorrect referent to the appropriate referent (i.e. switching from the 'some' referent to the 'all' referent in the *All* condition, or from the 'all' referent to the 'some' referent in *Some early* and *Some late* conditions), or switching from a correct referent to the incorrect referent (i.e. switching from the 'all' referent to the 'some' referent in the *All* condition, or the 'some' referent to the 'all' referent in *Some early* and *Some late* conditions) over seven consecutive 100ms time windows from the absolute offset of the quantifier (calculated individually for each item/condition combination; see Figure 4). The log-corrected average switching probabilities in each 100ms time window were then analysed using a 1-factor ANOVA, with condition (*All vs. Some early vs. Some late*) as the repeated measures factor. Significant differences in these analyses would indicate that switches between correct and incorrect referents were not uniformly launched across conditions. Significant effects were then followed up with pair-wise comparisons to establish the exact nature of effects.

Table 5 displays the statistical details of the effects, allowing generalization to participants (F_1) and items (F_2), for each time window of interest. Strength of association is reported in terms of partial eta-squared ($_p\eta^2$).

----------Figure 4 about here ----------

----------Table 5 about here ----------

Statistical analyses on the log-corrected switches to a correct target (OnTarget) revealed no significant differences between conditions in correcting an incorrect visual bias during the first 200ms following the quantifier offset. This is not surprising given that previous studies have suggested that it takes approximately 200ms to program and execute an eye movement (Hallett,

1986). However, from 200ms onwards we found reliable differences in the probability of switching to the correct target across the three conditions. Further analyses using Bonferoni pair-wise comparisons during these time windows indicated that the probability of switching from an incorrect target to a correct target was significantly lower in the *Some late* condition compared to both the *All* (200-300ms: ($t_1(29) = 2.22$, $p = 0.03$; $t_2(23) = 2.3$, $p = 0.03$); 300-400ms: ($t_1(29) = 2.28$, $p = 0.03$; $t_2(23) = 2.45$, $p = 0.02$); 400-500ms: ($t_1(29) = 2.91$, $p = 0.007$; $t_2(23) = 3.49$, $p = 0.002$); 500-600ms: ($t_1(29) = 3.7$, $p = 0.001$; $t_2(23) = 3.58$, $p = 0.002$); 600-700ms: ($t_1(29) = 3.99$, $p < 0.001$; $t_2(23) = 3.74$, $p = 0.001$)) and *Some early* (200-300ms: ($t_1(29) = 2.1$, $p = 0.04$; $t_2(23) = 2.33$, $p = 0.03$); 300-400ms: ($t_1(29) = 1.95$, $p = 0.06$; $t_2(23) = 3.02$, $p = 0.006$); 400-500ms: ($t_1(29) = 1.47$, $p = 0.15$; $t_2(23) = 2.86$, $p = 0.009$); 500-600ms: ($t_1(29) = 2.13$, $p = 0.04$; $t_2(23) = 2.11$, $p = 0.05$); 600-700ms: ($t_1(29) = 2.23$, $p = 0.03$; $t_2(23) = 2.14$, $p = 0.04$)) conditions. Switching rates did not differ significantly for *All* and *Some early* conditions in any of these time windows (All ts < 1.92). These results indicate that despite an early low-level bias to the 'all' referent, listeners were able to rapidly switch to the appropriate target within 300ms from auditory offset of the quantifier. Importantly, these baseline-corrected analyses suggest that this correct referent prediction occurs at the same rate regardless of whether the quantifier expression requires the listener to make an additional scalar implicature or not.

Statistical analyses of switches from a correct target to an incorrect target (OffTarget) revealed marginal differences between conditions during the first 300ms following the quantifier offset, and fully significant differences from 300ms onwards. Once again, these effects were followed-up with Bonferoni-corrected pair-wise comparisons, which showed that between 0ms and 400ms, the probability of switching to an incorrect target was significantly reduced following the quantifier "all" compared to "some" in the *Some late* condition (0-100ms: ($t_1(29) = 2.44$, $p = 0.02$; $t_2(23) = 1.76$, $p = 0.09$); 100-200ms: ($t_1(29) = 2.12$, $p = 0.04$; $t_2(23) = 1.94$, $p = 0.06$); 200-300ms: ($t_1(29) = 3.0$, $p = 0.005$; $t_2(23) = 2.71$, $p = 0.01$); 300-400ms: ($t_1(29) = 2.73$, $p = 0.01$; $t_2(23) = 3.05$, $p = 0.006$)), but did not differ from switches in the *Some early* condition

(all *t*s < 1.61). However, from 400ms onwards, the probability of switching to an incorrect target was significantly reduced following the quantifier "all" compared to "some" in both the *Some late* (400-500ms: ($t_1(29) = 3.97$, $p < 0.001$; $t_2(23) = 3.5$, $p = 0.002$); 500-600ms: ($t_1(29) = 4.63$, $p < 0.001$; $t_2(23) = 4.37$, $p < 0.001$); 600-700ms: ($t_1(29) = 2.6$, $p = 0.01$; $t_2(23) = 3.15$, $p = 0.004$)) and *Some early* (400-500ms: ($t_1(29) = 2.76$, $p = 0.01$; $t_2(23) = 3.93$, $p = 0.001$); 500-600ms: ($t_1(29) = 2.21$, $p = 0.04$; $t_2(23) = 3.06$, $p = 0.006$); 600-700ms: ($t_1(29) = 1.94$, $p = 0.06$; $t_2(23) = 2.41$, $p = 0.02$)) conditions. Interestingly, the probability of making an incorrect fixation did not differ between *Some early* and *Some late* conditions in any of these time regions (All *t*s < 0.9).

Thus in the *Some early* condition, it appears that despite participants' rapid ability to correct an initially incorrect fixation in favour of the quantity-appropriate referent, they do continue to be drawn to the 'all' referent. However, a comparison of the probability scores across the latter 500ms time windows in each trial type shows that listeners were more likely to correct an incorrect fixation than to direct their gaze from a correct referent to an incorrect one in both the *All* (62% vs. 25%) and *Some early* (59% vs. 39%) conditions. In contrast, the probability of switching gaze in the *Some late* condition only differed marginally between OnTarget and OffTarget trial types (35% vs. 42%).

Discussion

In this experiment we examined the timecourse with which anticipatory biases towards target areas of a visual display developed, based on interpreting a quantifier which does not call for any further inferences ('all'), versus a quantifier which invites an additional scalar implicature ('some' – 'but not all'). Our video items were intended to provide a rich context for participants, clearly focusing on the transfer of different quantities of objects to one of two potential locations. In this study we found that in the early regions of the relevant noun phrase ("pour some/ all"), a significant bias to the correct target emerged slightly earlier in the *all* condition compared to the *Some early* condition. More importantly, we found that participants were able to direct

anticipatory looks to the correct target in both the *All* and *Some-early* conditions long before the disambiguating expression ("limes"/ "oranges") was encountered, showing a correct bias in both conditions from the offset of the quantifier. Note that when equal quantities were transferred to both locations (*Some late*), participants did not show a bias to either referent at any point prior to linguistic disambiguation. In line with previous research in this area, we attributed the early small advantage for looks in the *All* condition relative to the *Some early* condition to be due at least in part to a low-level bias to the all referent in both *All* and *Some early* conditions (note that there is no all referent in the *Some late* condition). This explanation is supported by effects during the verb ("pour"), where a general all-referent bias emerged even *before* the onset of the quantifier. A follow-up analysis which examined the likelihood of a switch in gaze to the correct target showed that, from the onset of the quantifier, the likelihood of switching to the correct target was no different in the *All* and *Some early* conditions, whilst this was significantly greater than in the baseline *Some late* condition for both. Overall the results suggest that the utterance input had the same effect on fixation bias formation in the *All* and *Some early* conditions.

General Discussion

Scalar implicatures, such as the ones studied in this paper, are normally explained as implications that are derived via a reasoning process about what the speaker has said (and not said). In this paper we explored whether online access to scalar implicatures is delayed relative to access to an interpretation that does not involve quantity implicatures. In a visual world, 'look-and-listen' study, we found that participants' bias toward a relevant target emerged at the same rate when interpretation depended on quantity implicature ('some and not all') as when interpretation did not ('all').

This is the first study to compare the timecourse of access to scalar implicature relative to other information in on-line comprehension. Indeed it is the first timecourse study to look at access to conversational implicatures that enrich the meaning of an utterance beyond its semantic

interpretation. Our results suggest that there is no difference in the timecourse of access to scalar implicatures relative to aspects of 'what is said'. Previous research that addresses the timecourse of online access to Gricean inferences has focussed on cases where these inferences support the determination of the semantic interpretation of the utterance (or its presuppositions) because it is contributing to reference assignment. Thus we provide the first evidence that answers the research question originally set out in Huang and Snedeker (2009) about whether the semantic interpretation of an utterance is consulted prior to the derivation of conversational implicatures.

As to the discrepancy between the results in Huang and Snedeker (2009) and Grodner et al. (2010), we think that it is due in part to the fact that the pragmatic inferences serve the purpose of facilitating reference assignment, and in part to the presence of an additional 'numeral' condition. In Huang and Snedeker (2009), participants could either hear 'Point to the girl who has all of the soccer balls' or 'Point to the girl that has three of the soccer balls' as an instruction to point to the same referent. Similarly, they could hear either, 'the girl that has some of the socks' or 'the girl that has two of the socks'.[5] In Grodner et al. (2010) there is no numeral condition and so no alternative way of referring to the targets is made salient in the study. There is also no delay.

As mentioned above, Grodner et al. appeal to a pragmatic principle, 'Maximise Presupposition' to explain the effect. We think there are alternative possible explanations involving a principle of parsimony: that the speaker would not use a longer than necessary referring expression without a reason. Irrespective of which 'Gricean' principle we appeal to in explaining it, the effect relies on making an inference that if the speaker were referring to the all-

[5] Grodner et al. (2010) propose that two factors were responsible for the delay in Huang and Snedeker (2009). One was the presence of the numeral condition and the second concerned the phonological form of the quantifiers (saying 'summa' vs. 'some of'). It now seems clear that the delay in Huang and Snedeker's studies occurs when there is an additional 'numeral' condition. In subsequent unpublished work, Huang and Snedeker report a study that manipulates phonological form and presence of numerals as separate factors and finds that it is only the presence of numerals that leads to a delay. (See Huang & Snedeker, 'Some inferences still take time: Prosody, predictability, and the speed of scalar implicatures' CUNY 2010). Further evidence that the phonological form does not affect Gricean inference comes from the study reported in the current paper. Our speaker used the same un-elided form as in Huang and Snedeker (2009).

referent, they would use ‘all’ since that cues the participant to the correct referent unambiguously; and either way, the grounds for this inference are weakened in the presence of the items containing numerals. When numerals are present, there is a salient means of cueing the participant to the some-referent, this is to use the unambiguous ‘two’. Thus, on either kind of account, we can explain the interference effect of numerals in this task as being due to the fact that the pragmatic inference bears on the process of assigning reference to a definite expression. In our study where pragmatic inference plays a different role of enriching the overall interpretation, the presence of numeral items should not affect the timecourse of the implicature. This prediction is the subject of ongoing research.

One question left unanswered by this time-course research generally is what kinds of mechanisms are involved in the rapid access to Gricean inferences among other contextual inferences. If it was only lexically-triggered quantity implicatures that were accessed rapidly in online interpretation (as where ‘some’ triggers a ‘not all’ implication) then there might seem to be a relatively straightforward constraint-based account available. However, recent research from our lab has demonstrated that contextually specific quantity implicatures are accessed on-line, as where saying, ‘I enjoyed the dessert’ implies the speaker did not enjoy the main course (Breheny et al., 2010; under review). Moreover, it has been shown that even linguistically triggered quantity implicatures require contextual support, in terms of relevance to the conversational goal (Breheny et al., 2006). So, if we assume that rapid access is delivered by more-or-less automatic processes that are probabilistic, as in Constraint-Based or similar models, then the question arises about what aspects of context are being automatically monitored or inferred that can deliver these implicatures. One suggestion from the theoretical and computational literature is that an enriched common ground is constructed for all utterances, one that contains not only information necessary to determine what is said (e.g. to satisfy presuppositions) but also information about likely relevant and informative continuations. In some quarters, this aspect of common ground is described in terms of ‘Questions Under

Discussion' (see Roberts, 1996; Ginzberg, in press). Although very little online research has been conducted on this aspect of contextual processing, Tian, Breheny and Ferguson (2010) have shown that participants spontaneously accommodate something like a Question Under Discussion for even stand-alone, de-contextualised experimental items. The relevance of this research for timecourse research is clear: if comprehenders automatically compute not only context to determine the semantic interpretation of a sentence (e.g. the domain of quantification for 'some' or 'all') but also context to determine how the use of the sentence is likely to be relevant, then quantity implicatures that depend alternative ways the utterance could have been more relevant (or more informative) could be automatically triggered given that some alternatives are salient (see Geurts, 2010; van Rooij & Shulz, 2004; Spector, 2006, for details of how the relevant Quantity Implicature inferences would be warranted under these conditions). An interesting project for future pragmatic and psycholinguistic research would be to determine what exactly are the factors that lead to rapid access to implicatures.

References

Arai M., van Gompel R.P.G., & Scheepers, C. (2007). Priming ditransitive structures in comprehension. *Cognitive Psychology,* 54, *218-250*.

Altmann, G., & Steedman, M. (1988). Interaction with context during human sentence processing. *Cognition*, 38, 419-439.

Altmann, G.T.M., & Kamide, Y. (1999). Incremental interpretation at verbs: restricting the domain of subsequent reference. *Cognition*, 73, 247-264.

Altmann, G.T.M. & Kamide, Y. (2007). The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *Journal of Memory and Language,* 57, 502-518.

Altmann, G.T.M. & Kamide, Y. (2009). Discourse-mediation of the mapping between language and the visual world: Eye movements and mental representation. *Cognition*, 111, 55-71.

Bott, L., & Noveck, I.A. (2004). Some Utterances are Underinformative: The Onset and Time Course of Scalar Inferences, *Journal of Memory and Language*, 51(3), 437-457.

Breheny, R., Katsos, N., & Williams, J. (2006). Are Scalar Implicatures Generated by Default? *Cognition*, 100, 434-463.

Breheny, R., Ferguson, H.J., & Katsos, N. (2010) Taking the epistemic step. Paper presented at 23rd Annual CUNY Conference on Sentence Processing.

Breheny, R., Ferguson, H.J., & Katsos, N. (under review) Taking the epistemic step: Toward a theory of on-line access to conversational implicature.

Cooper, R.M. (1974). The control of eye fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology*, 6, 84-107.

Gazdar, G. (1979). *Pragmatics: Implicature, presupposition and logical form*. New York: Academic Press.

Geurts, B. (2010). *Quantity Implicature*. Cambridge: Cambridge University Press.

Ginzburg, J. (in press) *The Interactive Stance: Meaning for Conversation*. Oxford: Oxford University Press.

Grice, H.P. (1975). Logic and Conversation. In *The Logic of Grammar*, D. Davidson and G. Harman (eds), Encino, CA: Dickenson, 64-75.

Grodner, D.J., & Sedivy, J. (in press). The effects of speaker-specific information on pragmatic inferences. In N. J. Pearlmutter & E. Gibson (eds). The Processing and Acquisition of Reference. MIT Press: Cambridge, MA.

Grodner, D.J., Klein, N.M., Carbary, K.M., & Tanenhaus, M.K. (2010). "Some," and possibly all, scalar inferences are not delayed: Evidence for immediate pragmatic enrichment. *Cognition*.

Grodzinsky, Y., & Reinhart, T. (1993). The innateness of binding and co reference. *Linguistic Inquiry,* 24, 69-101.

Hallett, P.E. (1986). Eye movements. In K.B. Boff, L. Kaufman & J.P. Thomas (Eds.), *Handbook of Perception and Human Performance I: Sensory Processes and Perception*. pp 10-102. New York: Wiley.

Huang, Y.T., & Snedeker, J. (2009). Online Interpretation of Scalar Quantifiers: Insight into the Semantic-Pragmatics Interface. *Cognitive Psychology*, 58, 376-415.

Huang, Y.T., & Snedeker, J. (in press). 'Logic & Conversation' revisited: Evidence for a division between semantic and pragmatic content in real time language comprehension. To appear in *Language and Cognitive Processes*.

Heim, I. (1983). On the projection problem for presuppositions, in Barlow, M. and Flickinger, D. and Westcoat, M. (eds.), Second Annual West Coast Conference on Formal Linguistics, Stanford University, 114–126.

Heim, I. (1991) Articles and definiteness. Published in German as "Artikel und Definitheit," in A. v. Stechow and D. Wunderlich (eds.), *Semantics: An international handbook of contemporary research.* Berlin: de Gruyter.

Horn, L. (1972). On the semantic properties of the logical operators in English. Doctoral Dissertation, UCLA, Los Angeles, CA. Distributed by IULC, Indiana University, Bloomington, IN.

Horn, L. (1989). *A natural history of negation*. Chicago, IL: University of Chicago Press.

Levinson, S. (2000). *Presumptive Meanings*, MIT Press, Cambridge, Mass.

Neale, S. (1992). Paul Grice and the Philosophy of Language. *Linguistics & Philosophy* 15, 509-559.

Noveck, I., & Posada, A. (2003). Characterising the time course of an implicature. *Brain and Language*, 85, 203-210.

Reinhart, T. (1983). *Anaphora and semantic interpretation.* London: Croom Helm.

Roberts, C. (1996). Information Structure: Towards an integrated formal theory of pragmatics. In Jae Hak Yoon and Andreas Kathol (eds.) *OSUWPL Volume 49: Papers in Semantics.* The Ohio State University Department of Linguistics.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime reference guide*. Psychology Software Tools Inc, Pittsburgh, PA.

Sedivy, J. (2003). Pragmatic versus form-based accounts of referential contrast: Evidence for effects of informativity expectations. *Journal of Psycholinguistic Research*, 32, 3-23.

Sedivy, J.C., Tanenhaus, M.K., Chambers, C.G., & Carlson, G.N. (1999). Achieving incremental semantic interpretation through contextual representation. *Cognition,* 71*,* 109-147.

Spector, B. (2006). Aspects de la pragmatique des opérateurs logiques. Doctoral dissertation, University of Paris VII.

Sperber, D., & Wilson, D. (1986). *Relevance: Communication and cognition*. Oxford: Blackwell.

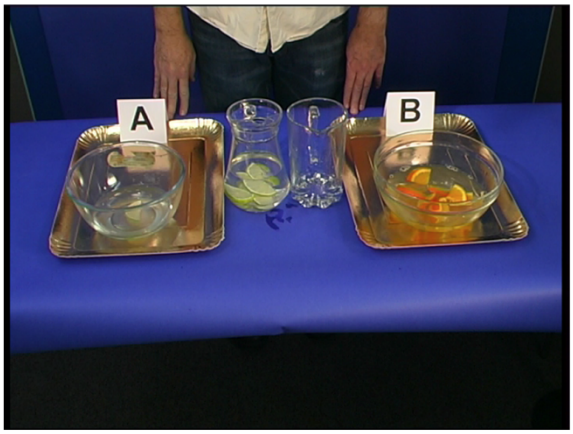
Tanenhaus, M., Spivey-Knowlton, M., Eberhard, K., Sedivy, J., 1995. Integration of visual and linguistic information during spoken language comprehension. *Science* 268, 1632–1634.

Tian, Y., Breheny, R. & Ferguson, H.J. (2010) We Simulate Negated Information: A Dynamic Pragmatic Account. *Quarterly Journal of Experimental Psychology.* 63(12), 2305-2312.

van Rooij, R., & Schulz, K. (2004). Exhaustive interpretation of complex sentences. *Journal of logic, language and information,* 13, 491-519.

Westerstahl, D. (1984). Determiners and context sets, in J. van Benthem and A. ter Meulen (eds.), *Generalized Quantifiers in Natural Language*, Foris, Dordrecht, 45-71.

Figure 1: Example visual stimulus. Participants heard sentence [1] or [2] (see above) whilst viewing picture [a], and sentence [3] whilst viewing picture [b].

[a] 

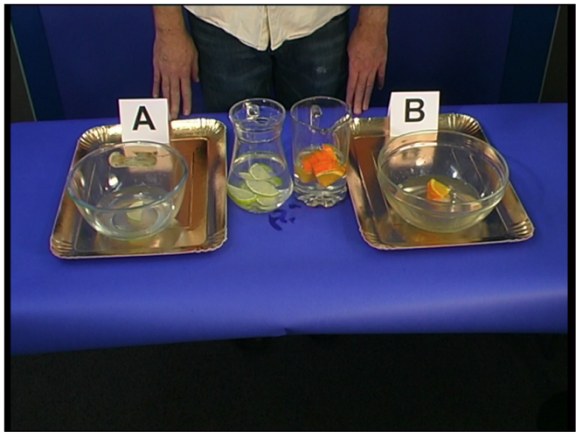
[b] 

Figure 2: Illustration of the experimental procedure.