In: Emar Maier, Corien Bary & Janneke Huitink, eds. (2005) "Proceedings of SuB9"
www.ru.nl/ncs/sub9

ARE SCALAR IMPLICATURES COMPUTED ONLINE?*

Gianluca Storto & Michael K. Tanenhaus
University of Rochester

storto@ling.rochester.edu; mtan@bcs.rochester.edu

Abstract

We adopt the *visual-world eye-tracking paradigm* to test the hypothesis that scalar implicatures are integrated very locally to the utterance of scalar terms. Focusing on the ⟨and,or⟩ scale, we show that early point-of-disambiguation effects similar to those triggered by the integration of the lexical meaning of and can be triggered by the integration of the *exhaustive* meaning of or. Some design issues and an independent interpretive asymmetry holding between and and or are discussed as possible explanations for remaining differences between the effects found in the two cases. We conclude that the exclusivity implicature that is associated to sentences containing or seems to be calculated online, rather locally to the utterance of the disjunction.

1 Introduction: Scalar implicatures

Since Horn (1972) the notion of conversational implicature proposed by Grice (see the papers collected in his 1989 book) has been put to use to explain certain interpretive differences between expressions in natural language and their counterparts in formal logic.

(1) a. Uli or Philippe asked questions after the talk.
$\leadsto$ Uli or Philippe asked questions after the talk, but not both.

b. Some students in the audience liked the talk.
$\leadsto$ Some students in the audience liked the talk, but not all.

The sentences in (1) seem to convey more than they would be expected to if the natural language disjunction or had the same meaning as the logical disjunction $\vee$, or if the quantificational determiner some was interpreted as the existential quantifier $\exists$. The intuitive meaning of the sentences in (1) imposes restrictions (the material underlined in the glosses) that go beyond the meaning of logical disjunction or existential quantification. Indeed, a logical formula like $P \vee Q$ is true if both disjuncts are, and a formula like $\exists x P(x)$ is true if the property P holds of all entities in the domain of quantification.

Horn proposes that the additional restrictions that seem to characterize sentences like those in (1) are not part of the lexical semantics of or and some, which does not differ

*We want to thank Ivano Caponigro, Greg Carlson, Christine Gunlogson, Carson Schütze, and Rachel Sussman for their input and comments on the design and results of this experimental work. Cassie Filios and Rebekka Puderbaugh helped us in testing subjects and coding data. Further useful input was provided by audiences at Sinn und Bedeutung IX, the UCLA Psychobabble seminar, a University of Maryland CNL Lab Lunch Talk, and the Tanenhaus Lab meetings in Rochester. This work was partially supported by a Ruth L. Kirschstein National Research Service Award (NIH grant T32 DC35-09) and by NIH grant HD-27206. Any remaining shortcomings in the form or content of this paper should be blamed on the first author.

from that of their logical counterparts. These additional restrictions associated with uses of or and some are implicatures, part of the pragmatic "overlay" that arises as a consequence of general rational cooperative behavior principles when natural language is used in conversational exchanges. In particular, the non-logical interpretation of or and some is due to a class of inferences that follow from Grice's first maxim of Quantity: "Make your contribution as informative as required for the current purpose of the exchange."

Horn points out that many expressions in natural language can be ordered into *linguistic scales*, i.e. sets of expressions of the same grammatical category that can be arranged in a linear order by degree of informativeness or semantic strength (2). If, as Grice argues, speakers routinely abide to conversational maxims like Quantity and take their interlocutors to do the same, use of a lower element on a linguistic scale implicates that the speaker is not in the position of using some higher (= stronger) element of the scale.

(2) *Linguistic Scales*
Ordered sets of expressions $\langle \alpha, \beta, \gamma, \ldots, \omega \rangle$, where by substituting α, β, γ, etc. in a sentential frame ϕ we obtain well-formed sentences $\phi(\alpha)$, $\phi(\beta)$, $\phi(\gamma)$, etc. such that $\phi(\alpha)$ asymmetrically entails $\phi(\beta)$, $\phi(\beta)$ asymmetrically entails $\phi(\gamma)$, etc.

In particular, uses of or or some, which share the property of being the weaker element in the linguistic scales ⟨and,or⟩ and ⟨all,some⟩, implicate that the speaker is not in the position of uttering the stronger statement containing and or all. Under the common assumption that the speaker's knowledge of the subject matter of the conversation is not incomplete, this *scalar implicature* conveys that the stronger sentence containing and or all is false. Hence uses of or convey but not both, and uses of some convey but not all.

As Bach (to appear, p.8) points out, "Grice did not intend his account of how implicatures are recognized as a psychological theory or even as a cognitive model. He intended it as a rational reconstruction. [...] He was not foolishly engaged in psychological speculation about the nature of or even the temporal sequence of the cognitive processes that implements that logic." Still, the misconception that implicatures in general, and scalar implicatures in particular, are late-arriving inferences which can be calculated only at later stages in the comprehension of a sentence is rather pervasive in the pragmatic literature.

Surprisingly, the issue has not received much attention in the experimental processing literature. Only a handful of recent contributions address the processing of scalar implicatures in adults (Breheny and Katsos 2003, Chierchia et al. 2003, Noveck and Posada 2003, Bott and Noveck 2004). But in general these works do not seem to focus on the actual timecourse of the computation of scalar implicatures: the processing of implicatures is probed offline, i.e. well after scalar items like or and some are presented to participants.

2 Experiment methodology and design

2.1 Experimental hypothesis

In this work, we aim at probing directly the timecourse of the computation of scalar implicatures, trying to determine whether this component of meaning is available at initial stages of processing or becomes available only at later stages. In particular, we focus on the ⟨and,or⟩ scale, testing the hypothesis that the exclusive component of the interpretation that is usually attributed to sentences containing a disjunction is computed and integrated very *locally* to the utterance of the disjunction or.

By saying that the exclusive meaning of or is calculated and integrated 'locally' to the utterance of the disjunction we mean the following. As an utterance unfolds, listeners try to integrate the information that can be extracted from what they have already heard into a (partial) representation of the content conveyed by the utterance. In particular, listeners access the information provided by the lexical meaning of words that they have heard. The integration of the lexical meaning of words is the paradigm of a very local process: as soon as a word is heard its lexical meaning (if known) becomes available and can be put to use. Our experimental hypothesis amounts to claiming that the implicated content that is normally associated with uses of the disjunction or does not differ much from lexical content. Like lexical content, the exclusive meaning of or should be "closely tied" to the utterance of this lexical item and become available as soon as the disjunction is heard.

In order to test this hypothesis, we adopt the so-called *visual-world eye-tracking experimental paradigm* (Tanenhaus et al. 1995). Within this paradigm subjects' gaze constitutes the dependent measure. Using a head-mounted eyetracker, gaze is tracked while subjects hear linguistic stimuli instructing them to perform actions on objects that are part of a "visual world" of reference—an array of actual objects or a display on a computer screen—which is concurrently presented to them.

The experimental paradigm builds on the observation that, when instructed to interact with an array of objects, subjects fixate the intended target of action significantly more often than other objects in the array (Eberhard et al. 1995). Thus, that a subject fixates one object in a given array significantly more often than the rest can be taken as an indication that the subject has uniquely identified the intended target of action. Of course, whether the intended target can be uniquely identified depends on both the nature of the instruction received and on the nature of the array of objects. In particular, if the interpretation of the instruction is determined in an incremental way, changes in the nature of the array of objects could potentially change the *point of disambiguation*, i.e. the point at which the instruction has provided sufficient information to uniquely identify the intended target of action. The nature of the array of objects presented to the subject, thus, can be manipulated in order to test specific hypotheses on the processing of linguistic stimuli.

The behavioral measure provided by the visual-world experimental paradigm is closely time-locked to the auditory stimulus. Subjects typically launch eye movements to the intended target of action within 500msec after the onset of the disambiguating word (Eberhard et al. 1995). Given that a latency of about 200msec occurs between the programming and the launch of eye movements (Matin et al. 1993), subjects initiate saccades to the target of action within 300msec from the onset of the disambiguating word.

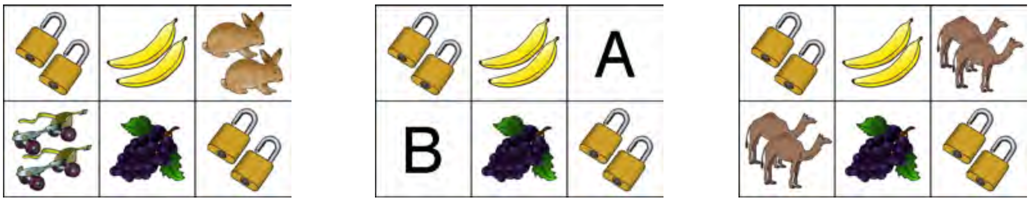
2.2 The logic of the experiment

It is probably easier to understand the logic of our experiment by looking first at a case in which only lexical meaning is at stake. Consider the meaning of the conjunction and (3). A conjunction of NPs in subject position denotes a function of type $\langle et, t\rangle$, which returns the value TRUE if applied to properties that hold of the denotation of both conjuncts. Essentially, understanding the meaning a conjunction of NPs in the subject position of a sentence amounts to knowing that the property denoted by the VP holds of both conjuncts.

(3) $[\![\textsf{and}]\!] = \lambda\mathcal{B}\lambda\mathcal{A}\lambda P[\mathcal{A}(P) \wedge \mathcal{B}(P)]$ $\rightsquigarrow$ look for a property shared by the conjuncts

If this information is integrated as soon as and is heard, we should be able to change the point of disambiguation in sentences containing a conjunction of NPs as subject by changing the number of properties shared by the objects denoted by the two conjuncts. In particular, if the only relevant properties are being next to certain or other types of objects, changing whether the squares marked with A and B in the display in Figure 1 (center) contain objects of the same type or of different types should have quite a dramatic effect on the point of disambiguation for sentence-instruction pairs like the one in (4).

(4) The bananas and the grapes are next to some locks. Please click on those locks.



one shared spatial property $\rightsquigarrow$ expected early disambiguation

two shared spatial properties $\rightsquigarrow$ expected late disambiguation

Figure 1: The logic of the experiment for the case of and

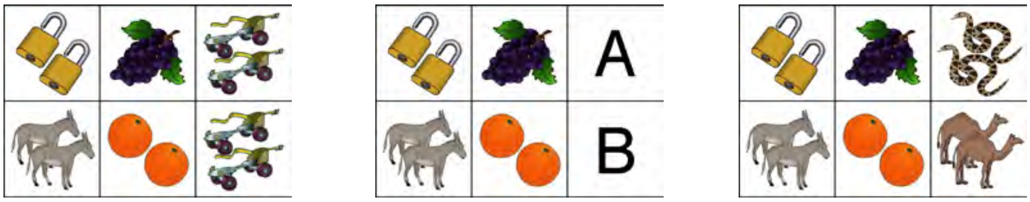
When the the objects in A and B are of different types (as in the display on the left in Figure 1), the bananas and the grapes share only the property of being next to some locks. If subject expect the follow-up instruction to ask them to perform some action on the objects mentioned in the VP of the first sentence, they should be able to uniquely identify the intended target of action already after having heard the conjunction and in the first sentence, i.e. before the intended target of action is mentioned at all. Conversely, when the objects in A and B are of the same type (as in the display on the right in Figure 1), the bananas and the grapes share both the property of being next to some lock and the property of being next to some camels. In this situation the integration of the meaning of and would not help subjects to identify the intended target of action, which could be distinguished from the other objects in the display only after being mentioned explicitly.

The same logic can be applied in investigating whether the exclusive meaning of or is integrated locally. Consider (5), where the exclusive component of the meaning of or is written directly into the lexical meaning of the disjunction. According to (5), a disjunction of NPs in subject position denotes a function of type $\langle et, t\rangle$ which returns the value TRUE if applied to properties that do not hold of both disjuncts, i.e. that differentiate the two. If this information is integrated as soon as or is heard, we should again be able to change the point of disambiguation of sentences containing a disjunction of NPs as subject by changing the number of properties shared by the objects denoted by the two disjuncts.

(5) $[\![\textsf{or}]\!] = \lambda\mathcal{B}\lambda\mathcal{A}\lambda P[\mathcal{A}(P) \vee \mathcal{B}(P)\ \underline{\&\ \neg(\mathcal{A}(P) \wedge \mathcal{B}(P))}]$
$\rightsquigarrow$ look for a property that distinguishes the disjuncts

Once again, we can expect that changing whether the squares marked with A and B in the display in Figure 2 (center) contain objects of the same type or of different types should affect the point of disambiguation for sentence-instruction pairs like the one in (6).

(6) The grapes or the oranges are next to some locks. Please click on those locks.



one shared spatial property ⇝ expected <u>early</u> disambiguation

no shared spatial properties ⇝ expected <u>late</u> disambiguation

Figure 2: The logic of the experiment for the case of or

Leaving empty properties aside, the function denoted by exclusive or is the complement of the function denoted by and: every nonempty set that is mapped to TRUE by 'NP_1 or NP_2' is mapped to FALSE by 'NP_1 and NP_2', and viceversa. This means that if the objects in A and B are of the same type (as in the display on the left in Figure 2), subjects should be led to disregard them as possible targets of action already after having heard the disjunction or in the first sentence. While this does not by itself uniquely identify the intended target of action—until hearing locks two alternatives remain open—the integration of the exclusive meaning of or should be reflected in an increase in looks to the two remaining potential targets of action. Conversely, when the objects in A and B are of different types (as in the display on the right in Figure 2) the integration of the exclusive meaning of or would not help subjects in "narrowing down" the set of potential targets of action, and looks should be more equally distributed among the four possible alternatives.

Crucially, however, only the local integration of the *exclusive* meaning of or is expected to distinguish between the display on the left and the display on the right in Figure 2. If subjects integrate the exclusive component of the meaning of the disjunction only at a later point in the processing of sentences like (6), or should be initially given the same *inclusive* interpretation as the logical disjunction $\vee$, an interpretation that would not exclude the roller skates as potential intended targets of action in the display on the left.

Thus, this experimental design allows us to probe the participants' interpretation of or without setting up an explicit verification task, where subjects would be asked to consciously evaluate the interpretation(s) licensed by a sentence containing the disjunction. Behavioral data from such tasks likely conflate and confound the participants' processing of the linguistic stimuli with the verification strategy adopted to perform the task. In addition, an explicit verification task might encourage subjects to consider from the start interpretations that are not immediately considered in the normal processing of sentences.

2.3 A summary: Experimental conditions and predictions

Before addressing the design of the experiment, let us summarize the various experimental conditions and the predictions that follow from our experimental hypothesis.

We first investigate the effects of the integration of lexical content using the paradigm detailed for the case of and above. This preliminary step is essential in order to test that our experimental methodology works. Indeed, results like those in (Eberhard et al. 1995) concern primarily the effects of the integration of the meaning of open-class content words—adjectives, in particular—rather than more "functional" close-class words like and or or. Furthermore, the methodology that we adopt departs slightly from the basic visual-world paradigm: we are interested in tracking the participants' gaze while they hear a sentence that describes the visual display, rather than while being instructed to perform an action. Still, in our design subjects should process the first sentence in order to identify which objects the action requested by the following instruction should be performed on. We expect the basic results of the visual-world paradigm to be replicated within this setting.

For the case of and we consider two conditions (see Figure 3): an early disambiguation condition (AE), where we expect the integration of the meaning of the conjunction to help subjects in identifying the intended target of action already before it is mentioned in the VP of the sentence, and a late disambiguation condition (AL), where we expect the intended target of action to remain ambiguous until it is explicitly mentioned.

AND LATE AND EARLY OR LATE OR EARLY OR INCLUSIVE

Figure 3: The 5 experimental conditions

Then, we test whether the effects that we expect to find in the case of and can be reproduced for the case of or, using the same logic detailed above. As in the case of and, we have two basic conditions for or: an early disambiguation condition (OE), and a late disambiguation condition (OL). In addition, we introduce a third condition (OI), which resembles the early disambiguation condition in that the same kind of displays are used, but which differs from it in that the two identical objects in the display are mentioned in the auditory stimuli as intended targets of action. Items of this sort, in which the disjunction in the first sentence must be interpreted as inclusive, are needed in order to avoid biasing subjects towards an exclusive interpretation of or. But these items are not mere fillers. Given our hypothesis that subjects should initially be driven away from shared properties by the exclusive meaning of or, we might expect a further disambiguation delay in the OI condition, similar to the garden-path effects discussed in syntactic processing literature.

3 The experiment

3.1 Methods

3.1.1 Materials

The actual displays in the experimental materials consisted of 3×3 square grids containing 9 (pairs of) objects. Adding a third row in the display was necessary to ensure that subjects had to process the first sentence in order to correctly perform the action requested by the follow-up instruction. If the simpler displays shown in the previous section had been used, subjects could have easily adopted a heuristic—"click on objects of the type mentioned in the VP of the first sentence"—that would have allowed them to perform correctly the requested action without actually paying attention to the meaning of the conjunction/disjunction in subject position. With the more complex display, we can ensure that subjects process the first sentence by varying whether the follow-up instruction requires them to click on the objects mentioned in the VP of the first sentence or on some other objects of the same type: the third row contains an additional pair of objects of this type. Consider the two alternative sentence-instruction sequences in (7) with respect to the display in Figure 4: those and other in the instructions can be interpreted only with respect to the information conveyed by the the first sentence.[1]

(7) The bananas and the grapes are next to some locks.
  a. Please click on those locks.
  b. Please click on some other locks.

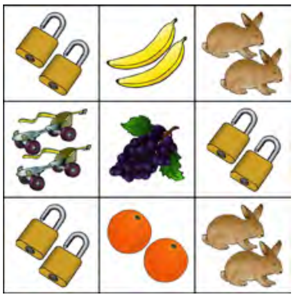


Figure 4: An example of the full 3×3 grid displayed in a trial

The pairs of objects used in the displays were constructed using images from the color Snodgrass picture set (Rossion and Pourtois 2004). The central column—containing pictures of bananas, grapes, and oranges—remained constant across all displays. The other two columns contained pictures chosen among the eight pairs of objects in Figure 5.

We chose to consider as experimental items only displays where the objects denoted by the subject of the first sentence are in contiguous rows. For each of the AL, AE, OL, and OE conditions 4 items were created. The OI condition consisted of 8 items, in order to offset the exclusive interpretation of or required by the OL and OE conditions. In half of the experimental items the third row appears above the two relevant rows, and in the other half it appears below them. 12 filler items were created that are essentially identical to experimental items but for the fact that the two rows referred to by the subject of the first sentence are not contiguous. In addition 12 more filler items were created in which sentences like (8) are used to describe displays like those used in experimental items.

[1]As pointed out by Carson Schütze (p.c.), the instruction in (b) is potentially confusing in the case of sentences containing a conjunction in subject position as in (7). The instruction could be interpreted as requiring to click on the two pairs of locks in the top two rows or on just one of these pairs. Both types of actions were considered as correct in analyzing the data. This potential source of confusion is removed in the follow-up experiments that we are currently running, as are the potential problems raised by the possibility of interpreting the indefinite in object position as scoping over the subject in the first sentence.

Figure 5: The 8 pairs of objects used in the experiment

Finally, 64 filler items were created for which sentences like those in (9) are used as descriptions of the visual display. Altogether, the set of test items consisted of 112 items.

(8) Some locks are next to the bananas and/or the grapes.

(9) a. The bananas are next to some locks.
b. Some locks are next to the bananas.

Care was paid in balancing this set as evenly as possible. All eight objects appeared as intended targets the same number of times,[2] and overall all objects occurred equally often in the set of test items. Four different lists of experimental items were created. The 8 objects in Figure 5 were divided in two sets in order to ensure that different objects appeared as intended targets in the 4 AL/OL vs. the 4 AE/OE items, and one factor of difference between the lists was which set was used in which condition. Balancing the distribution of the intended targets of action and the remaining "alternative" objects among the 4 possible cells available in the grid would have required to create 8 items per condition. We chose to divide the possible alternative layouts in two sets, and have the choice between these two sets be the second factor of difference between the four lists of experimental items. Finally, the order of mention of objects in the conjunctions and disjunctions in the first sentence in the auditory stimuli was balanced too, as was whether the follow-up instruction designated as target of action objects in the rows referred to by the subject of the first sentence or the relevant object in the third row. These auditory stimuli were recorded as whole sentences spoken with normal intonation by a female native speaker of English.

3.1.2 Participants

Participants were sixteen (16) male and female undergraduates from the University of Rochester Department of Brain and Cognitive Sciences subject pool, all of whom were paid for their participation.[3] All participants were native speakers of North American English with normal or corrected to normal vision and no hearing impairments. The participants were equally distributed among the four lists of experimental materials.

[2]One note about terminology. Since in our analysis we consider only looks to the two rows referred to by the subject of the first sentence, from now on we will use the term 'intended targets' to refer only to objects of the type mentioned in the VP that appear in these rows.

[3]Fixation data from 6 additional participants were not analyzed: $<$90% correctness in performing the requested action was taken to indicate that participants were not attending to the experimental task.

3.1.3 Procedure

The experimental materials were presented using an Apple eMac computer equipped with a 17-inch monitor (1024×768 pixel resolution) and external stereo loudspeakers. During the experiment participants were seated about 30 inches away from the computer monitor.

Each trial began with the presentation of a number at the center of the blank screen. After 500msec the number disappeared and a 3×3 square grid (768×768 pixels) containing 9 (pairs of) objects was displayed. After 3 seconds the first sentence—e.g. The bananas and the grapes are next to some locks—was played, followed by a 300msec pause and then by the follow-up instruction—e.g. Please click on those locks. After performing the requested action, participants pressed the spacebar to go to the next trial.

Before testing proper, subjects were were presented with four practice trials in order to familiarize with the task to be performed. Practice trials differed from the trials in the testing phase in that objects other than those in Figure 5 were used in the visual displays, and in that subjects received explicit feedback on their performance in the follow-up task.

A PsyScope script (Cohen et al. 1993) controlled the presentation of the stimuli and recorded the subjects' performance in the follow-up task. The 112 items in the testing phase were presented in random order in one block (subjects were allowed to take breaks between trials). A run of the experiment took on average about 30 minutes.

Participants' eye movements were monitored using an ISCAN EC-501 head-mounted eyetracker. An eye camera provides an infrared image of the eye and tracks its position by analyzing the positions of the center of the pupil and the first Purkinje reflection. A scene camera is aligned with the participant's line of sight, providing a context with respect to which eye position data is localized. Output from the scene camera, along with a superimposed crosshair marking point of gaze, and the audio signal, were recorded for the whole experiment using a Sony Digital-8 professional editing VCR. Audio and video signals were synchronized; the recording camera samples at a rate of 30 frames per second and each video frame was stamped with a time code. Eye-tracker calibration was monitored and adjusted as necessary by the experimenter between trials.

For experimental trials, a frame-by-frame editing VCR was used to identify looks to the 9 cells on the screen. Coders did not know which cells contained intended targets of fixation, nor did they hear the auditory stimuli that were played in each trial. Subsequent automatic post-processing of the coded data identified the objects fixated in each trial.

3.2 The case of and

3.2.1 Results

The results are expressed here as fixation proportions over time, pooling across all trials falling into a given condition. The graphs in Figure 6 show the proportion of fixations to *target* vs. *alternative* in the AL and AE conditions. For each frame (recorded on the x-axis), looks to target vs. alternative (recorded on the y-axis) are calculated as follows. Taking the sentence in (4) as paradigm, target looks is the average amount of looks to the two cells containing locks in the two "relevant" rows of the display divided by the total number of looks to the screen in that frame, and alternative looks is the average amount of looks to the two "other" cells in the relevant rows divided by the same number. A frame was coded as containing a look to a cell if either the participant was fixating the cell or the eye was in transit to that cell during a saccadic eye movement. The vertical bars on the graphs mark the frames corresponding on average to the beginning of the conjunction, the noun in the second conjunct, the verb in the VP, the noun in the object NP, the 300msec pause, and the follow-up instruction in the auditory stimuli.

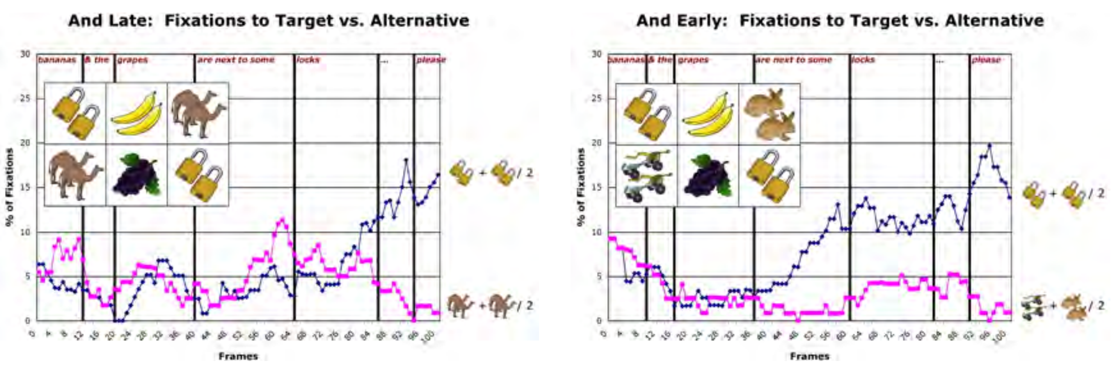


Figure 6: Fixation timelines (and)

Visual inspection of the graphs reveals that, as expected, participants converged on fixating the target much earlier in the AE condition than in the AL condition. In AL looks to target (blue line) vs. alternative (purple line) diverge only after the beginning of the object noun, but the two diverge already after the second conjunct is played in the AE condition.

To more closely investigate disambiguation, we divided each auditory stimulus into time windows, corresponding to the seven regions delimited by vertical bars in the above graphs. The length of these windows varies on a per-item basis due to differences in the duration of the recorded stimuli. The start point and end point of each window were offset 200ms (6 frames) to account for the approximate amount of time needed to plan and launch a saccade based on incoming auditory information.[4] For the first four regions following the conjunction—corresponding to (i) the noun in the second conjunct, (ii) the VP minus the object noun, and (iii) the object noun in the first sentence, and (iv) the pause between the first sentence and the follow-up instruction—we conducted an omnibus ANOVA with subjects as a repeated measure. Event (2nd Coord NP, Verb, Object NP, Pause), Condition type (Late, Early), and Object fixated (Target, Alternative) were within-subjects factors. The dependent measure was the average proportion of fixations in each time window.

The ANOVA reveals a significant Event×Condition×Object interaction ($F(3,45)=3.64$, $p=0.0196$). Planned comparisons show that the effect is due to differences between the AL and AE conditions in the participants' preference for fixating the target vs. alternative objects. In AL participants do not display a preference for the target until the pause between the first sentence and the follow-up instruction. On the other hand, in AE participants prefer to fixate the target already while they hear the verb of the first sentence ($F(1,45)=16.26$, $p=0.0002$). Figure 7 shows the difference between the mean fixation to target and the mean fixation to alternative in the AL and AE conditions for the four time windows; values for which this difference is statistically significant are circled.

[4]E.g. to ensure that auditory information about the object noun can influence eye movements in the corresponding window, the start point for the analysis window needs to be 6 frames after the onset of locks.

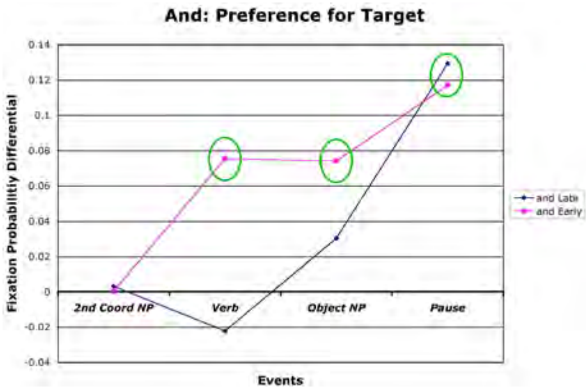


Figure 7: Preference for target (and)

3.2.2 Discussion

The results suggest that participants in this study were able to access and integrate the lexical meaning of and very locally to the utterance of the conjunction and use this information to guide the further processing of the sentence. In particular, note that disambiguation of the target occurs immediately after hearing the second conjunct in the AE condition. As soon as the two relevant rows are identified, participants can use the information provided by the subject of the first sentence to uniquely identify the target, well before the target itself is explicitly mentioned in the auditory stimulus.[5]

This shows that the experimental methodology adopted in this study is appropriate to the task. Within the design of the experiment, the behavioral measure provided by the visual-world paradigm can detect local effects of the integration of lexical semantic information. We can thus turn to testing our experimental hypothesis: the implicated content of or—i.e. its exhaustive interpretation—should trigger similar local disambiguation effects.

3.3 The case of or

3.3.1 Results

The graphs in Figure 8 show the proportion of fixations to target vs. alternative in the OL and OE conditions. The way of computing looks to target and alternative is a little different in this case. Consider the OE condition first, assuming that subjects hear (6) as auditory stimulus. Under our experimental hypothesis we expect subjects to look away from the two pairs of roller skates, which thus constitute the alternative. But what about the target? One option would be to consider the other two cells in the two relevant rows as target. However, this is appropriate only until the word locks begins to be played: after that, we expect subjects to concentrate on the locks alone. Thus, for the OE condition we decided to compare the proportion of looks to the single cell containing the intended target to the

[5]When we presented our results, people in the audience voiced the concern that the early POD effect found in the AE condition might be due to properties of the visual stimuli. In AE subjects might prefer looking towards the two "alike" objects rather than towards the two "different" ones. This alternative account for the effect found in AE fails in light of the findings for the OE condition, which is visually indistinguishable from AE but does not seem to induce a comparable preference for the two alike objects.

average proportion of looks to the two cells containing identical objects. Items in the OL condition were constructed so that for each item in the OE condition a corresponding item existed that contained the intended target in the same cell, but replaced the two identical alternative objects with two different objects. Looks to alternative in the OL condition were calculated by averaging looks to the two cells containing these different objects.[6]

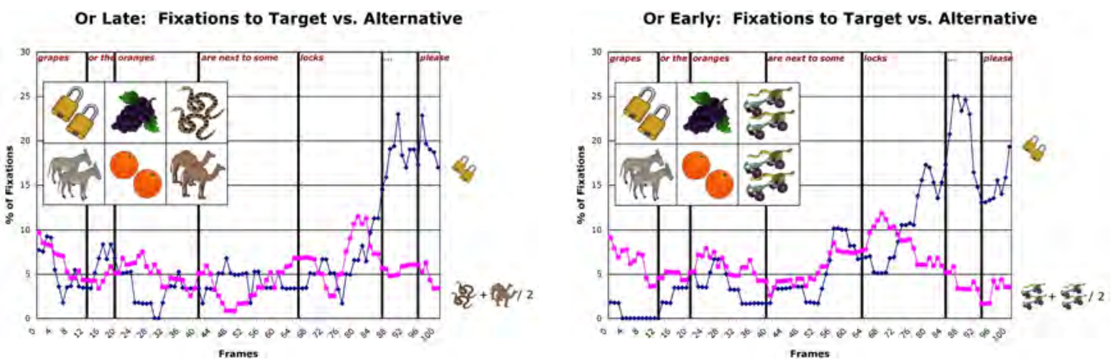


Figure 8: Fixation timelines (or)

Visual inspection of the graphs reveals that participants converged on fixating the target earlier in the OE condition than in the OL condition. In OL looks to target vs. alternative clearly diverge only after the end of the first sentence; in OE the two diverge while the object noun is being heard. The expected early POD effect is thus found in the OE condition, but this effect seems to be delayed with respect to the effect found in the case of and.

In order to better understand the results, we conducted an omnibus ANOVA with subjects as a repeated measure on the first four windows following the disjunction—corresponding to (i) the noun in the second disjunct, (ii) the VP minus the object noun, and (iii) the object noun in the first sentence, and (iv) the pause between the first sentence and the follow-up instruction. Event (2nd Coord NP, Verb, Object NP, Pause), Condition type (Late, Early, Inclusive), and Object fixated (Target, Alternative) were within-subjects factors. The dependent measure was the average proportion of fixations in each time window.

In this case, the ANOVA does not reveal a significant Event$\times$Condition$\times$Object interaction. However, planned comparisons show that the OE condition differs from the OL and OI conditions with respect to the participants' preference for fixating the target. In OL and OI participants do not display a preference for the target until the pause between the first sentence and the follow-up instruction. On the other hand, in OE participants prefer to fixate the target already while they hear the object noun in the VP of the first sentence ($F(1,90)=10.713$, $p=0.0015$). Figure 9 shows the difference between the mean fixation to target and the mean fixation to alternative in the OL, OE, and OI conditions for the four time windows; values for which this difference is statistically significant are circled.

3.3.2 Discussion

The early POD effect found in the OE condition suggests that participants were able to locally use the exclusive meaning of or to guide the further processing of the sentence and restrict the set of possible targets. Notice that while disambiguation of the target occurs

[6]Looks to target vs. alternative in the OI condition were calculated exactly as in the AE condition.

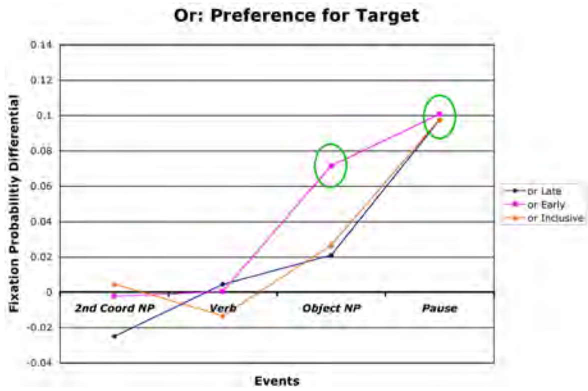


Figure 9: Preference for target (or)

in OE in the time window corresponding to the object noun (e.g. locks), the effect cannot be due to the explicit mention of the target. As shown by the results for the AL, OL and OI conditions, effects of the integration of the lexical meaning of the object noun can be detected only in the time window corresponding to the pause after the first sentence.

This argues that the exclusive interpretation of or is available at an early stage in the processing of the first sentence. Indeed, were subjects to initially interpret or as the logical disjunction $\vee$, differences in their behavior in the OE condition vs. the OL and OI conditions would not be expected until after the integration of the meaning of the object noun.

4 General discussion

Two main results follow from our experiment. The first result is that the methodology that we devised seems to allow for investigating the meaning of functional words like and or or without setting up an explicit verification task. The data collected using this methodology are less likely to conflate or confound the effects of the integration of the meaning of these expressions with those due to strategies adopted by the participants.

The second result is that we find evidence that the exclusive component of the meaning of or is integrated (and thus calculated) online. Our experimental analysis was led by the hypothesis that the exclusive component of the interpretation that is normally associated with sentences containing the disjunction or is calculated very locally to the utterance of this lexical item. It appears that or is given an exclusive interpretation already before the sentence containing the disjunction has been processed in its entirety, which clearly undermines the "extreme" alternative to our experimental hypothesis that many authors seem to have implicitly attributed to Grice.[7] The exclusive interpretation of or seems to be available to participants at a point where the "literal meaning" of the sentence containing the disjunction cannot be calculated because the sentence has not been heard in its entirety.

At the same time, the most extreme version of our locality hypothesis does not seem to be upheld by the results either. Participants in our experiment do not seem to use the information provided by the exclusive meaning of or as early as they use the information provided by the lexical semantics of and. The early POD effect attested in the OE con-

[7]"If Grice is right [. . .] you need to know the literal meaning or sense of a sentence before you can calculate its implicatures in a context [. . .]" (Levinson 1983, p.117).

dition occurs later, and is thus less local than the effect found in the AE condition.[8] This provides a potential argument against the hypothesis that the exclusive component of the meaning of or becomes available as soon as the disjunction is heard.

For the purpose of this paper, we would like to hold on drawing the latter conclusion. Our reluctance in abandoning the strong version of our experimental hypothesis is motivated by the observation that the asymmetry found between the effects in the case of and vs. or might be due to independent issues arising from specific properties of our experiment.

One potential problem follows from our choice of investigating the $\langle$and,or$\rangle$ scale in the first place. An independent formal asymmetry holds between the two elements in this scale: while a conjunction of NPs can denote both in e and $\langle et, t\rangle$, a disjunction of NPs is inherently non-referential. This asymmetry might be playing an unwanted role in our experiment because the experimental task in the visual-world paradigm is essentially a referential one: subjects are implicitly asked to determine an interpretation for the auditory stimulus with respect to the referential domain provided by the visual display. It is thus possible that the delayed effect found in the OE condition indicates a delayed integration of the whole meaning of or, and not just of its exclusive component. That is, or could be interpreted as exhaustive as soon as it is heard, but its meaning be of a type that—unlike the meaning of and—cannot be used right away in the visual-world setting.

A second problem is that we unwillingly introduced a strong bias in the experimental materials that militates against the effects that we expected to find in the OE and OI conditions. Consider again the visual display in Figure 4. A subject faced with a display of this type who always chose to concentrate on the cells containing the three identical objects would be 75% correct in guessing the identity of the target, without paying any attention to the nature of the coordination in the subject of the first sentence. Such a strong bias is likely to have been unconsciously picked up by participants, with the result of undermining both the early POD effect in OE and the expected disambiguation delay in OI.

In ongoing follow-up experiments we address these potential confounds, improving the experimental design and extending the scope of investigation to the $\langle$all,some$\rangle$ scale, where an asymmetry similar to that occurring in the case of $\langle$and,or$\rangle$ does not arise.[9]

5 Conclusions

Our experiment provides initial evidence that the exclusive meaning of or is integrated locally to the utterance of the disjunction, and can guide the further processing of the sentence containing it. Like other types of linguistic information, scalar implicatures seem to be computed and integrated online, as part of the incremental processing of a sentence.

As a parting note, we want to explicitly state that we do not intend to draw conclusions bearing directly on the current theoretical debate on the nature of scalar implicatures from the provisional results of our experimental investigation. Like Grice, most contenders in

[8] Furthermore, we do not find the related expected further disambiguation delay in the OI condition.

[9] Within the revised design, we plan to address relevant questions that have been raised by the audience at *Sinn und Bedeutung IX*. In particular, experimental items are blocked in order to test whether the participants' behavior changes with exposure to the task; and versions of the experiments are planned in which no conjunction items are presented, in order to test the hypothesis that the local exclusive interpretation of the disjunction might be triggered by an implicit comparison to sentences containing a conjunction.

the theoretical arena aim at an appropriate rational reconstruction of the logic underlying the derivation of implicatures and of the types of information involved in it, and do not commit to hypotheses concerning the use of this knowledge that can be straightforwardly translated into behavioral predictions. Still, we think that these and further experimental results can contribute to the debate by defining empirical requirements that a psychologically realistic analysis of scalar implicatures should be able to meet at no additional cost.

References

Bach, K.: to appear, The top 10 misconceptions about implicature, *in* B. Birner and G. Ward (eds), *A Festschrift for Larry Horn*, John Benjamins, Amsterdam.

Bott, L. and 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. and Katsos, N.: 2003, Are generalised conversational implicatures generated on-line by default?, Paper presented at the *Workshop on Polarity, Scalar Phenomena, and Implicatures*, University of Milan, Bicocca.

Chierchia, G., Clifton, Jr., C. and Frazier, L.: 2003, When basic meanings are (not) enough: Processing scalar implicatures during adult language comprehension, Manuscript, University of Milan, Bicocca and University of Massachusetts, Amherst.

Cohen, J. D., MacWhinney, B., Flatt, M. and Provost, J.: 1993, PsyScope: A new graphic interactive environment for designing psychology experiments, *Behavioral Research Methods, Instruments, and Computers* **25**(2), 257–271.

Eberhard, K. M., Spivey-Knowlton, M. J., Sedivy, J. C. and Tanenhaus, M. K.: 1995, Eye movements as a window into real-time spoken language comprehension in natural contexts, *Journal of Psycholinguistic Research* **24**, 409–436.

Horn, L.: 1972, *On the Semantic Properties of Logical Operators in English*, PhD thesis, University of California, Los Angeles.

Levinson, S. C.: 1983, *Pragmatics*, Cambridge University Press, Cambridge.

Matin, E., Shao, K. and Boff, K.: 1993, Saccadic overhead: Information processing time with and without saccades, *Perception & Psychophysics* **53**, 372–380.

Noveck, I. A. and Posada, A.: 2003, Characterizing the time course of an implicature: An evoked potential study, *Brain and Language* **85**, 203–210.

Rossion, B. and Pourtois, G.: 2004, Revisiting Snodgrass and Vanderwart's object pictorial set: The role of surface detail in basic-level object recognition, *Perception* **33**(2), 217–236.

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