The Time Course of Scalar Implicature

Lewis Bott & Ira A. Noveck

Abstract for the International Workshop "Where Semantics Meets Pragmatics"
Michigan State University, July, 11-13, 2003

In paper presents an experimental investigation into a class of inference known as *scalar implicatures*. These arise when a less-than-maximally-informative utterance is taken to imply the denial of the more informative proposition (or else to imply a lack of knowledge concerning the more informative one). Consider the following dialogues:

1) Peter: Are Cheryl and Tony coming for dinner?
Jill: Cheryl or Tony is coming.

2) John: Did you get to meet all of my friends?
Robyn : Some of them.

In (1), Jill's statement can be taken to mean that *not both* Cheryl and Tony are coming for dinner and, in (2), that Robyn did not meet all of John's friends. These interpretations are the result of scalar implicatures, which we will describe in detail below. Before we do so, note that the responses in each case are -- from a strictly logical point of view -- compatible with the questioner's stronger expectation; if Jill knows that both Cheryl and Tony are coming, her reply is still true and if in fact Robyn did meet all of John's friends, she also spoke truthfully. *Or* is logically compatible with *and* and *some* is logically compatible with *all*.

Such inferences were first classified by Paul Grice as *generalized implicatures*, as he aimed to reconcile logical terms with their non-logical meanings. Grice, who was especially concerned by propositional connectives, reflected on those inferences accompanying logical terms that become, through conversational contexts, part of the speaker's overall meaning. In one prime example, he described how the disjunction *or* has a weak sense, which is compatible with formal logic's (the inclusive-or), but as benefiting from a stronger sense (*but not both*) through conversational uses (which would make the disjunction exclusive). What the disjunction *says*, he argued, is compatible with the weaker sense, but through conversational principles it often *means* the stronger one. Any modern account of the way logical terms are understood in context would not be complete without considering these implicatures.

Scalar implicatures have been discussed at length in the linguistic-pragmatic literature as it has greatly expanded on Grice's original insights. In what follows, we present descriptions of scalar implicature from the point of view of two post-Gricean pragmatic theories that aimed to elaborate on Grice but are now in often in conflict. One approach is often referred to as *neo-Gricean* (Horn, 1973; Levinson 1983, 2000) and the other is known as Relevance Theory (Sperber & Wilson, 1985 1996; Carston, 2002). We will discuss each account of scalar implicature in turn while ultimately focusing on existential quantifiers.

According to neo-Griceans like (Horn, 1973) and Levinson (1983, 2000), the scalars described in (1) and (2) above are paradigmatic cases of implicature that work on *terms* that are relatively weak. The speaker's choice of a weak term implies the rejection of a stronger term from the same scale. For example, the terms *some* and *all* may be viewed as part of a scale (<some, all>), where *all* constitutes the more informative element of the scale (since *all p* entails *some p*). In the event that a speaker chooses to utter *some* the hearer will take it as suggesting that the speaker has no evidence that the stronger element in the scale holds (i.e. it is not that case that *all* holds). Neo-Griceans believe that the implicature to deny the stronger term in the scale arises automatically as a "default" or "preferred " meaning. For example, the default interpretation of *some* is *some but not all*. This implicature can be cancelled, but only in certain contexts subsequent to the production of the scalar.[1]

The other account comes from Relevance Theory (Sperber & Wilson, 1985 1996), which assumes that an utterance can be inferentially enriched in order to better appreciate the speaker's intention, but this is not done on specific words as a first step to arrive at a default meaning. According to Relevance Theory, a scalar is but one example of pragmatic implicatures that arise when a speaker intends and expects a hearer to draw an interpretation of an utterance that is relevant enough. How far the hearer goes in processing an utterance's meaning is governed by principles concerning effect and effort; namely, listeners try to gain as many effects as possible for the least effort.

A non-enriched interpretation of a scalar term (the one that more closely coincides with the word's meaning) could very well lead to a satisfying interpretation of this term in an utterance. Consider *Some monkeys like bananas*. This utterance with an interpretation of *Some* that remains in its weaker form (this can be glossed as *Some and possibly all monkeys like bananas*) can suffice for the hearer and not require further pragmatic enrichment. In contrast, the potential to derive a scalar implicature comes into play when an addressee applies relevance more stringently. A scalar implicature could well be drawn by a hearer in an effort to make an utterance, for example, more informative. Common implicatures like scalars are implicatures that optionally play a role in such enrichment; they are not

[1] Levinson does specify contexts in which the scalar implicature could be, in effect, preempted from occurring. One example (entailment) is when prior context blocks the scalar implicature because it would be inconsistent. Consider the following as prior context (from Levinson, 2000, p. 50) : *A Saudi Prince has just bought Harrod's;* this would block the production of a scalar implicature (*Some but not all*) in *Some Saudi princes must be pretty wealthy*. The existentially quantified statements that we will investigate here remain unembedded and thus should not preempt scalar implicatures according to Levinson.

steadfastly linked to the words that could prompt them. If a scalar does arrive in a context that renders an underinformative utterance more informative, it ought (all things being equal) to be linked with extra effort.

In this paper, we present two experiments to test between the neo-Gricean and the Relevance theory explanation of scalar implicature. Both of these experiments ask participants to judge the veracity of category sentences involving quantifiers. For example, a participant might see the sentence “All elephants are mammals” and would then have to judge whether the statement was true or false. The sentences of most interest are sentences of the form *Some X are Y*, where, in fact, all X are Y. An example of this type of sentence would be “Some monkeys are mammals”. This type of sentence will be considered false if the participant makes the implicature (so the sentences becomes ‘Some but not all monkeys are mammals’), but true if the participant makes the strictly logical interpretation of the term *some* (‘Some and possibly all monkeys are mammals’). We refer to “false” responses to this type of sentence as *pragmatic* and “true” responses as *logical*. We also present a variety of control sentences involving other quantifiers and other category relationships, such as “All birds are trout” (see Table 1). The sentences discussed earlier of the form *Some X are Y* will be referred to as Underinformative sentences to distinguish them from the Control sentences.

Experiment 1

According to neo-Griceans, a Pragmatic response to Underinformative sentences should be quicker than a Logical response. This is because they consider the default interpretation of *some* to be *some but not all* and this is the first interpretation to be considered by the participant. Similarly, the Logical response should require a relatively long response time because the *some but not all* interpretation must be cancelled before the *some and possibly all* interpretation is made. In contrast, Relevance Theory predicts that processing effort is required to make the pragmatic enrichment of scalar terms such as *some*. This means that more time should be required to make the implicature *some but not all* and consequently to respond Pragmatically to the Underinformative sentences.

Method

Participants. Thirty-two undergraduates from the Université de Lyon 2, who were either volunteers or presented with a small gift worth about 4 Euros, participated in this study.

Stimuli and Design. Participants saw six types of sentences. These are shown in Table 1, together with an example of each. Participants saw 9 examples of each type of sentence, making a total of 54 sentences. For each participant, the experimental sentences were generated randomly from a base of 6 categories and 9 exemplars from each of these categories (see Appendix). This randomization procedure was adopted to eliminate, or at least minimize, any unwanted effects of frequency or typicality on the reaction times.

Table 1
Examples of the Sentence Types used in Experiments 1-3

Reference	Example sentence	Appropriate Response
T1	Some elephants are mammals	?
T2	Some mammals are elephants	T
T3	Some elephants are insects	F
T4	All elephants are mammals	T
T5	All mammals are elephants	F
T6	All elephants are insects	F

Note. T1 sentences are the underinformative sentences referred to in the text. The question mark in the Correct Response column indicates that T1 can be considered true or false depending on whether the participant draws the implicature or not.

Procedure. Participants were placed in front of a computer and told that they would see sentences presented on the screen. The only instructions participants were given was to respond ‘True’ if they thought the sentence on the screen was true, or ‘False’ if they believed the sentence to be false. Participants were not told whether their responses were correct or incorrect, i.e. there was no feedback.

Results

Data treatment. Outliers were considered to be responses made in less than 0.5 seconds or more than 10 seconds. This resulted in 12 % trials being removed from the data set. Incorrect answers to the Control sentences were eliminated from the analysis involving reaction times. This resulted in an additional 10% of the responses being removed.

Analysis of choice proportions. The nine individual trials for each sentence type were pooled, producing a set of six means per participant. For the 5 control sentences, participants were largely in agreement in choosing true or false responses: Correct responses for T2 through T6 ranged from 87% to 98%.%. As demonstrated elsewhere (Noveck 2001), responses to Underinformative sentences prompt a high degree of bivocality - 61% of responses were pragmatic interpretations.

Analysis of reaction times. In order to assess whether a logical response was made more quickly than a pragmatic response, we divided each participant’s answers to Underinformative sentences into Logical and Pragmatic and then found the mean reaction time for these two groups. This gave us a within-participant measure of the change in reaction time for response type. However, 9 participants were excluded from the analysis because they responded to all trials using a single type of response – either all Logical (2) or all Pragmatic (7). Figure 1 shows the mean reaction times for the six sentence types, with T1 divided into Logical and Pragmatic responses. Pragmatic responses for T1 sentences take longer than Logical responses. This trend was confirmed by performing a paired t-test between the average time taken to respond Pragmatically and the average time taken to respond Logically ($t_1(22)=2.07$, $p = 0.05$; $t_2(5) = 4.7$, $p = 0.0054$). Further analysis demonstrated that Pragmatic responses to T1 sentence required more time than to process than responses to any of the control sentences (all p_1's < 0.05; all p_2's < 0.06), while Logic responses to T1 sentences required the same amount of time as responses to the majority of control sentences (all p_1's > 0.13; all p_2's > 0.25).

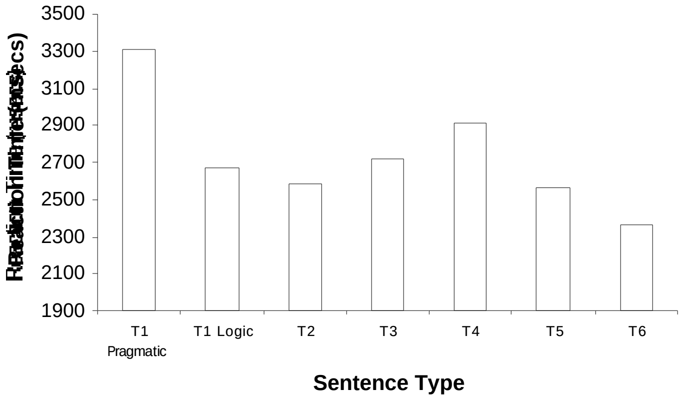

Figure 1.

Discussion

The main finding here is that mean reaction times were longer when participants responded pragmatically to the Underinformative sentences than when they responded logically. Furthermore, pragmatic responses to the Underinformative sentences were slower than responses to all of the control sentences. Collectively, our findings provide evidence against default implicatures because there is no indication that participants require more time to arrive at a true response for the Underinformative sentences. All indications point to the opposite being true: Logical responses to Underinformative sentences are indistinguishable from responses to control sentences while Pragmatic responses to Underinformative are significantly slower.

Although our experiments provide evidence against the idea that scalar implicatures become available as part of a default interpretation, they do not necessarily provide evidence in direct support of the alternative presented here, the Relevance theory explanation. (Moreover, a theorist in the original Gricean tradition could take the results from Experiments 1 as supportive to Grice's theory because the data point to a distinction between an initial semantic interpretation and a pragmatic one.) Our goal in the next experiment is to test directly predictions from Relevance theory concerning the processing of scalar implicature.

Experiment 2

According to Relevance theory, implicatures are neither automatic nor arrive by default. Rather, they are cognitive effects that are determined by the situation and, if they do manifest themselves, ought to appear costly compared to the very same sentences that do not prompt the implicature. In Relevance terminology, all other things being equal, the manifestation of an effect (i.e. the implicature) ought to vary as a function of the cognitive effort required. If an addressee (in this case, a participant) has many resources available, the effect ought to be more likely to occur. However, if cognitive resources are rendered limited, one ought to expect fewer implicatures. Experiment 2 tests this prediction directly by varying the cognitive resources made available to participants. The experiment follows the general procedure of Experiment 1, in that participants are asked to judge the veracity of categorical statements. The crucial manipulation is that the time available for the response is varied; in one condition participants have a relatively long time to respond (referred to as the Long condition), while in the other they have a short time to respond (the Short condition). By requiring participants to respond quickly in one condition, we intend to limit the cognitive resources they have at their disposal. Note that it is only the time to *respond* which is manipulated; participants are presented with the words one word at a time and at the same rate in both conditions, thus there is no possibility that participants in the Short condition spend less time reading the sentences than those in the Long condition.

Relevance Theory would predict fewer implicatures when participants' resources are limited. It is expected that they would be more likely to respond with a quick "True" response when they have less time than when they have more. If one wanted to make predictions based on default interpretations, *some* should be interpreted to mean *some but not all* more often in the short condition than in the long condition (or at least there should be no difference between the two conditions).

Method

Participants. Forty-five participants were used in the study. Participants were either volunteers or were presented with a small gift worth about 4 Euros.

Stimuli and design. Participants again had to respond true or false to 54 category statements, generated in the same way as in Experiment 1. The new independent variable was the time that participants were given to respond to the statement, referred to as the Lag. The Lag was a between participant variable which could be either a short time (900 ms) after the presentation of the final word, or a long time (3000 ms). The dependent measure was the proportion of true responses within the time lag.

Procedure. The instructions for both conditions were similar to those of the previous experiment. In both Long and Short conditions, participants were instructed that if they took too long to respond they would see a message informing them of this. In the Short condition, speed of response was emphasized and participants were told that they would have to respond in less than half a second. The trial by trial procedure was similar to that of Experiment 1 until the participant made their response. After the response, the participant was told whether they were 'in time' or 'too slow'. In the Short condition they were 'in time' if they responded in less than 900 ms, whereas in the Long condition the limit was 3000 ms.

Results

Data treatment. Responses that were outside the allotted time lag for each condition were removed from the analysis. Thus, responses were removed if they had an associated reaction time of more than 900 ms in the Short condition and more than 3000 ms in the long condition. This resulted in a total of 12 % eliminated from the Short condition and 0.7% from Long condition. There appeared to be a uniform distribution of removed responses across the different sentence types.

Analysis. Table 2 shows the rates of True responses for all six sentence types. The rate of correct performance among the control sentences either improves (T3 - T6) or remains constant (T2) with added response time. This trend is shown in the last column of Table 2 which, for control sentences, indicates the increase in proportion correct with added response time. In contrast, responses to the Underinformative sentences were less consistent with added time available. This change was such that there were more Logical responses in the Short condition than in the long condition: 72% True in the Short Lag condition and 56% True in the Long lag condition. This trend is in line with predictions made by Relevance theory. A t-test revealed that there were significantly more Logical responses in the Short Lag condition than in the long Lag condition ($t_1(43) = 2.43$, $p = 0.038$; $t_2(5) = 6.6$, $p < 0.005$), although no other sentence types showed a reliable effect after multiple comparisons had been taken into account.

Table 2
Summary of results for Experiment 2

Sentence	Example	Short Lag	Long lag	Response difference
T1	Some elephants are mammals	0.72 (0.053)	0.56 (0.095)	-0.16
T2	Some mammals are elephants	0.79 (0.021)	0.79 (0.038)	0.00
T3	Some elephants are insects	0.12 (0.012)	0.09 (0.007)	+0.03
T4	All elephants are mammals	0.75 (0.027)	0.82 (0.024)	+0.07
T5	All mammals are elephants	0.25 (0.061)	0.16 (0.022)	+0.09
T6	All elephants are insects	0.19 (0.017)	0.12 (0.011)	+0.07

Note. The Short lag and Long lag columns contain the proportion of True responses for each condition. The final column refers to the increase in consistency of responses with added response time. For control sentences this equates to the increase in proportion correct with more time, while for the T1 sentences the figure is the Long condition True response minus the Short condition True response.

Discussion

This experiment manipulated the time available to participants as they were making a categorization judgements. We found that when a short period of time was available for participants to respond, they were more likely to respond "True" to T1 sentences. This strongly implies that they were less likely to derive the implicature when they were under time pressure than when they were relatively pressure-free.

The control sentences provide a context in which to appreciate the differences found among the T1 statements. They showed that performance in the Short Lag condition was quite good overall. In fact, the 72% who responded "True" in T1 represented the lowest rate of consistent responses in the Short condition. All of the control sentences in both the Short and Long lag conditions were answered correctly at rates that were above chance levels. For the control sentences, correct performance increased with added time.

The results of Experiment 2 provide further evidence against the neo-Gricean claim of default generation of the implicature. Furthermore, we feel that this experiment confirms a very specific prediction of Relevance Theory - that a reduction in the cognitive resources available will reduce the likelihood that an implicature will be made.

General Discussion

The experiments presented in this paper were designed to compare the neo-Gricean and the Relevance Theory account of scalar implicature. Experiment 1 demonstrated that a pragmatic interpretation of a sentence involving a scalar implicature took longer than a logical interpretation. These results lend doubt to the neo-Gricean claim that the default treatment of *some* is *some but not all*. Experiment 2 presented a more direct test of the Relevance account. Cognitive resources were manipulated (by way of time available for responding) to see whether fewer resources were linked with fewer implicatures. In the Experiment, those who had less time to respond to Underinformative items (900 msecs), responded using a logical interpretation at rates that were above chance levels. Meanwhile, they also answered the control items correctly at rates that were even higher. As this account would predict, when resources were made more available by way of increased time (3 seconds), it coincides with more implicature production and, thus, higher rates of pragmatic interpretations. Taken together, these findings indicate that people initially employ the weak, linguistically encoded meaning of *some* before employing stronger senses, arguably derived by a scalar implicature.

Until now, we have concentrated on theoretical linguistic-pragmatic accounts for the way scalar implicatures are drawn out of *some*. Here we consider a psychological possibility, which is that the error rates and slowdowns related to pragmatic readings of *some* results from the nature of the *some but not all* proposition itself. This explanation places the weight of the slowdown not on drawing the implicature *per se*, but on the work required to determine the veracity of a proposition with the implicature embedded within it. There are two ways in which the *some but not all* proposition is more complex than, say, *some but possibly all*. One is that such a proposition gives rise to a narrower set of true circumstances; thus determining whether or not a statement is true requires more careful assessments. The other is that negation, as is often the case, adds costs to processing (Just & Carpenter, 1971; Clark & Chase, 1972; although see Lea & Mulligan, 2002). Both of these suggestions are worthwhile descriptions of the cause of implicature-related slowdowns and worth further study. However, neither of these is inconsistent with Relevance theory's account, which makes the original counterintuitive prediction that the pragmatically enriched interpretation requires effort. Both of the above suggestions would have to have recourse to Relevance theory in order to explain its *a priori* predictions and results from Experiment 2, which showed how reduced resources lead to fewer implicatures.

In summary, this work largely validates distinctions made by Grice nearly a half-century ago by showing that what a term like *some* initially *says* is consistent with its logical reading. What it is understood to *mean* depends on the listener drawing further implicatures. This study focused on the manner in which implicatures are drawn. They do not appear to be general and automatic as neo-Griceans like Levinson claim. Rather, as outlined by Relevance Theory, implicatures occur in particular situations as an addressee makes an effort to render an utterance more informative.

References

Carston, R. (2002). *Thoughts and Utterances*. Oxford: Blackwell.

Chierchia, G., Crain, S., Guasti, M., Gualmini, A., & Meroni, L. (2001). The acquisition of disjunction: evidence for a grammatical view of scalar implicature. In A. H. J. Do (Ed.), *BUCLD Proceedings* (Vol. 25, pp. 157-168). Somerville, MA: Cascadilla Press.

Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior, 12*(4), 335-359.

Clark, H. H., & Chase, W. G. (1972). On the process of comparing sentences against pictures. *Cognitive Psychology, 3*, 472-517.

Evans, J. S. B., & Newstead, S. E. (1980). A study of disjunctive reasoning. *Psychological Research, 41*(4), 373-388.

Fillenbaum, S. (1974). Or: Some uses. *Journal of Experimental Psychology, 103*(5), 913-921.

Grice, H. P. (1989). *Studies in the way of words*. Cambridge, MA: Harvard University Press.

Horn, L. R. (1973). *Greek Grice: A brief survey of proto-conversational rules in the History of Logic*. Proceedings of the Proceedings of the Ninth Regional Meeting of the Chicago Linguistic Society. 205-214. Chicago

Just, M. A., & Carpenter, P. A. (1971). Comprehension of negation with quantification. *Journal of Verbal Learning and Verbal Behavior, 10*, 244-253.

Lea, R. B. (1995). On-line evidence for elaborative logical inferences in text. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 21(6)*, 1469-1482.

Lea, R. B., & Mulligan, E. J. (2002). The effect of negation on deductive inferences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*(2), 303-317.

Lea, R. B., O'Brien, D. P., Fisch, S. M., Noveck, I. A., & et al. (1990). Predicting propositional logic inferences in text comprehension. *Journal of Memory and Language, 29*(3), 361-387.

Levinson, S. (1983). *Pragmatics*. Cambridge: Cambridge University Press.

Levinson, S. (2000). *Presumptive meanings: The theory of generalized conversational implicature.* Cambridge, MA: MIT Press.

Moxey, L. M., Sanford, A. J., & Dawydiak, E. J. (2001). Denials as controllers of negative quantifier focus. *Journal of Memory and Language, 44*(3), 427-442.

Newstead, S. E. (1995). Gricean implicatures and syllogistic reasoning. *Journal of Memory & Language, 34(5)*, 644-664.

Noveck, I. A. (2001). When children are more logical than adults: Experimental investigations of scalar implicature. *Cognition, 78*(2), 165-188.

Noveck, I. A., Chierchia, G., Chevaux, F., Guelminger, R., & Sylvestre, E. (2002). Linguistic-pragmatic factors in interpreting disjunctions. *Thinking and Reasoning, 8*(4), 297-326.

Papafragou, A., & Musolino, J. (2003). Scalar implicatures: experiments at the semantics-pragmatics interface. *Cognition, 86*(3), 253-282.

Paris, S. G. (1973). Comprehension of language connectives and propositional logical relationships. *Journal of Experimental Child Psychology, 16(2)*, 278-291.

Paterson, K. B., Sanford, A. J., Moxey, L. M., & Dawydiak, E. (1998). Quantifier polarity and referential focus during reading. *Journal of Memory and Language, 39*(2), 290-306.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*(10), 437-442.

Rips, L. J. (1975). Quantification and semantic memory. *Cognitive Psychology, 7*(3), 307-340.
Sanford, A. J., Moxey, L. M., & Paterson, K. B. (1996). Attentional focusing with quantifiers in production and comprehension. *Memory and Cognition, 24*(2), 144-155.
Sperber, D., & Wilson, D. (1985 1996). *Relevance: Communication and Cognition.* Oxford: Basil Blackwell.