Distinguishing the SAID from the IMPLICATED Using a Novel Experimental Paradigm

Meredith Larson[*], Ryan Doran[*], Yaron McNabb[†], Rachel Baker[*], Matthew Berends[*], Alex Djalali[*], and Gregory Ward[*]

Northwestern University[*] and University of Chicago[†]

1. Introduction

That there is a theoretical distinction between context-dependent and context-independent aspects of utterance interpretation has become a standard assumption in current theories of meaning; however, how and where to draw this distinction has been the subject of considerable debate. In this paper, we investigate the extent to which speakers, when provided with consistent criteria, are able to systematically distinguish between WHAT IS SAID and WHAT IS IMPLICATED in an experimental setting. Our study aims to build upon previous empirical studies by (a) focussing on truth-conditional meaning (as opposed to focussing solely on implicature), (b) examining a wider range of implicature types, and (c) providing a clear set of criteria to guide participants in the experimental task. Our findings suggest that participants are generally able to distinguish implicatures from the truth-conditional meaning of an utterance and that the availability of an objective third-person perspective enhances this ability.[1]

2. Theoretical Background

Grice introduced his notion of implicature to account for the ability of speakers to communicate more than the mere compositional meaning of the speakers' words (Grice 1967, 1987). Implicatures are those aspects of speaker-intended meaning that go beyond the truth-conditional meaning of the utterance (WHAT IS SAID). When uttering a sentence, one says what the words conventionally mean and implicates additional intended meaning, which together constitute WHAT IS COMMUNICATED. For Grice, all conversational implicatures are CALCULABLE (according to conversational maxims), CANCELLABLE, and generally NON-DETACHABLE. However, our exclusive concern in this paper will be GENERALIZED CONVERSATIONAL IMPLICATURES (GCIs), which are distinct from PARTICULARIZED CONVERSATIONAL IMPLICATURES (PCIs) in that GCIs arise under normal circumstances unless there is something in the context that prevents them from doing so. This leads Levinson (2000) to argue that GCIs constitute default interpretations. For the purposes of this study, the important feature of conversational implicatures is that they are cancellable. Since conversational implicatures are by definition not part of truth-conditional meaning, a speaker can deny, or cancel, an implicature without also denying the truth of WHAT IS SAID.

While acknowledging Grice's insight that speakers are able to communicate more information pragmatically via utterance meaning than what is encoded semantically in sentence meaning, post-Griceans dispute the nature and extent of the category of implicature (Sperber & Wilson 1986, Récanati 1997, Carston 2002, *inter alia*). Their approach takes issue with the notion of saying as used by Grice, as well as with whether any implicatures arise as default interpretations. Instead, post-Griceans argue that many purported examples of GCIs actually intrude upon truth-conditional meaning and should not be considered implicatures at all. Rather, GCIs are taken to be pragmatically-determined aspects of speaker meaning, constitutive of utterance meaning and part of WHAT IS SAID. Thus, if one is unable to cancel an implicature without also denying the truth of WHAT IS SAID, then the purported implicature is part of WHAT IS SAID.

3. Previous Empirical Studies

Post-Gricean approaches, and Relevance Theory in particular, have paid considerable attention to questions of pragmatic processing and have thus lent themselves well to empirical investigation. On-line

[1] We would like to thank Matt Goldrick for his help in designing the experiments and running the statistics reported on in this paper. We are also grateful for the helpful comments from the members of the audience at the LSA 2007 Annual Meeting and the 2007 Experimental Pragmatics (XPRAG 2007) Conference where earlier versions of this work were presented.

studies, such as those that measure reading times, provide useful insights into how speakers process certain pragmatic aspects of meaning. However, they do not directly address questions about where to draw the line between WHAT IS SAID and WHAT IS IMPLICATED. These studies focus on processing rather than the products of processing, as Bezuidenhout and Cutting (2002: 439) rightly note. The first study to directly address the question of whether speakers do in fact systematically distinguish between WHAT IS SAID and WHAT IS IMPLICATED is Gibbs and Moise 1997. This study included several experiments designed to test, among other things, whether participants distinguished between "minimal" (i.e. WHAT IS SAID) and "enriched" (i.e. GCI-inclusive) interpretations and further whether participants' judgments were sensitive to training on Grice's distinction between saying and implicating. Throughout the experiments in this study, the instructions made explicit use of the term "said". Participants' responses in the first experiment significantly favoured the enriched interpretations as best reflecting what they thought each sentence said. The second experiment replicated the first but included additional instructions explaining Grice's distinction by means of examples. Even with the additional instructions, participants still overwhelmingly chose the enriched interpretations as best reflecting what they thought the stimuli said.

These results have been taken to lend support to post-Gricean views, as participants seemed to be observing the saying/implicating distinction, yet they nonetheless identified examples of GCIs as part of WHAT IS SAID. However, Nicolle and Clark (1999) were unable to fully replicate these results. In an effort to refine the methodology of Gibbs and Moise 1997, one of Nicolle and Clark's experiments in this series utilized three conditions in which participants were asked to select paraphrases that best reflected "what the speaker said", "what the speaker's words meant", and "what the speaker wanted to communicate". Surprisingly, participants in each of these conditions preferred the PCI interpretation over one with an enriched GCI-inclusive reading. Since these results do not support the idea that speakers are distinguishing between saying and implicating, Nicolle and Clark addressed this issue in a subsequent experiment by having participants choose between GCI-inclusive and PCI paraphrases. Their third experiment significantly improved upon the methodology of Gibbs and Moise 1997 by asking participants to interpret utterances according to the technical use of saying as explained at the outset of the experiment. While Gibbs and Moise (1997) provided participants with an explanation of saying and implicating, they did not ask participants to answer in accordance with the technical usage of saying, allowing them to employ their own notion. The results of Nicolle and Clark's final experiment did show that participants could distinguish said and implicature readings. However, the stimuli for this third experiment included only a GCI-enriched paraphrase for the WHAT IS SAID choice, rather than paraphrases corresponding to Grice's notion of WHAT IS SAID. Thus, this experiment does not provide data that speak to the question of whether speakers consider GCIs to be part of WHAT IS SAID, although it does suggest that when they are provided with consistent criteria, speakers can distinguish PCIs from GCIs.

These studies are generally taken to constitute support for post-Gricean views in general and Relevance Theory in particular. However, when it comes to the question of the truth-conditional status of GCIs and whether they are to be included as part of WHAT IS SAID, these studies do not adequately address the issue in three respects. First, the methodology employed in these experiments either did not ask participants to use the saying/implicating distinction to guide their responses or did so but did not present a choice between the WHAT IS SAID and GCI-inclusive interpretations. Second, these experiments did not attempt to investigate systematically the wide variety of examples that have been variously classified as GCIs in the literature. Gibbs and Moise (1997), Nicolle and Clark (1999), and Bezuidenhout and Cutting (2002, experiment 2) use at most six different types of GCI sentences.[2] These six types represent only a small sample of purported GCIs, primarily those covered by Levinson's I-principle (2000) or Horn's R-principle (1993). Third, these studies have not addressed the question of the relationship between GCIs and WHAT IS SAID from the perspective of the possible truth-conditional contribution of GCIs. Previous experiments have focussed on speakers' judgments concerning the meaning of "what is said", by using

[2] The six types of GCIs used in these studies were: cardinals, quantifiers (classified as "scalars" in Gibbs and Moise 1997), possessives, time-distance, temporal relations, and perfectives.

this very term as well as other technical vocabulary in the instructions. However, use of such technical terminology without any explanation is problematic because it is not clear whether participants were understanding the instructions as intended. Moreover, even when participants in these studies were provided with an explanation of technical terms, there were no independent measures included as part of the experiment to ensure that participants had in fact understood the task.[3]

4. Methods

In the current study, we address three questions left unanswered in the previous literature and experimental work on GCIs. First, can speakers access a level of meaning corresponding to the Gricean notion of WHAT IS SAID that is exclusive of GCIs? Previous experiments suggest that speakers cannot, but those experiments do not directly address the issue of truth-conditional meaning. Second, do different types of GCIs behave similarly with respect to truth-conditional meaning? Stimuli from previous experiments were limited to a small subset of the types of GCIs found in the literature, and there have been no attempts to investigate systematically whether these additional types have any bearing on whether GCIs are distinguishable from WHAT IS SAID. Third, can we improve upon previous methodologies? In previous experiments—even when the instructions introduced participants to Grice's original distinctions—it is unclear what informed participants' judgments, and whether and to what degree they were guided by the instructions particular to the experiment.

4.1 Overview of the experiment For each stimulus item, participants were asked to perform three tasks: read a short conversation, make a forced-choice truth evaluation about a sentence in the conversation, and rate their confidence in their evaluation. The truth-evaluation task tested whether a licensed GCI was incorporated into truth-conditional meaning without the assumption that participants had mastered technical terminology such as WHAT IS SAID. The experiment consisted of one within-subjects factor consisting of four stimulus types and one between-subjects factor consisting of three different instruction conditions. The same stimuli were used in each condition. The four stimulus types will be discussed in Section 4.2, and the three different instruction conditions will be discussed in Section 4.3.

4.2 Materials All experimental, filler, and control stimuli had the same design, which consisted of a short conversation of one or two exchanges between two speakers, Irene and Sam. For each exchange, Irene would ask Sam a question to which Sam would respond. Each conversation was followed by a fact relevant to the conversation. The underlined portion of Sam's response was the target whose truth participants were instructed to evaluate with respect to the provided fact. We created 88 of these short conversations: 28 controls (14 entailments and 14 contradictions), 16 filler items, and 44 target items containing GCIs. There were 4 examples of each of the 4 types of fillers and each of the 11 types of GCIs. Many of the GCI examples were adapted from examples gleaned from previous studies. The division into four stimulus types (contradictions, entailments, fillers, and GCIs) enabled us to test whether GCIs as a group patterned differently from the other stimulus types. The filler and target items will be discussed in Sections 4.2.2 and 4.2.3, respectively.

4.2.1 Controls: Entailments and contradictions In order to ensure that participants were attending to the truth-evaluation task, we used 14 entailment and 14 contradiction items. An example of a contradiction item is provided in (1).

(1) *Contradiction item*

Irene: When did Robert's great-uncle Jake die?
Sam: <u>He died in 1963.</u>
FACT: Robert's great-uncle Jake died in 1957.

For the contradiction items, the underlined portion of Sam's reply is false given the FACT. For the

[3] An orthodox Gricean could, after all, maintain that the results of the previous experiments which attempted to train participants on the notions of saying and implicating show that the participants did not master the distinction described in the instructions.

entailment items, the underlined statement is entailed by the FACT, as in (2) below.

(2) *Entailment item*

Irene: Tell me about Sophia's car.

Sam: <u>She owns a gray Chevy.</u>

FACT: The car Sophia owns is a gray Chevy.

Participants were predicted to judge Sam's statement to be false in the contradiction items and true in the entailment items. Data from these stimuli were used to determine whether participants were attending to and understanding the task.

4.2.2 Fillers: Necessary contextual elements Grice acknowledged that identifying WHAT IS SAID in some cases requires a pragmatic contribution in order to arrive at truth-evaluable meaning, such as the disambiguation of lexical items, the resolution of reference, the fixing of indexicals, and the recovery of ellipsis. While such examples constitute necessary pragmatic elements of WHAT IS SAID, Grice thought them to be few in number and different from implicatures in that they must first be interpreted in order to establish any truth-conditional meaning. In the absence of any umbrella term in the literature to describe these elements, we will refer to them as NECESSARY CONTEXTUAL ELEMENTS of meaning (NCEs) for the purposes of our study. We used four types of NCEs in the experiment: deictics, ellipses, indexicals, and pronoun resolutions. An example of each is given in (3).

(3) *Necessary contextual elements (NCEs)*

a. <u>Deictic</u>

Irene: What shoes are you wearing to dinner?

Sam: <u>I'm going to wear these shoes.</u>

FACT: Sam has decided to wear the shoes in the upstairs closet, not the ones he is currently putting on.

b. <u>Ellipsis</u>

Irene: What did everyone eat?

Sam: <u>Robert ate apples, oranges, and pears, and so did Melissa.</u>

FACT: Robert ate apples, oranges, and pears, and Melissa ate only apples and oranges.

c. <u>Indexical</u>

Irene: Did you have your annual dentist appointment yet?

Sam: Yes. <u>I went yesterday.</u>

FACT: Sam went to his annual dentist appointment on the same day as this conversation.

d. <u>Pronoun resolution</u>

Irene: I haven't seen that coat I gave you for Christmas... And what did you do with the sweater I gave you?

Sam: <u>I hung it in the closet.</u>

FACT: Sam hung the coat Irene gave him in the closet, and he put the sweater from Irene in his dresser drawer.

For the NCE stimuli, the most natural interpretation of Sam's response is false given the FACT. All stimuli were constructed so that the NCE element of Sam's utterance would be truth-conditionally inconsistent with the information contained in the FACT. For example, in (3c), if Sam went to his dentist appointment on the day of the conversation, then the most natural interpretation of his utterance must be false.

4.2.3 Experimental items: GCIs Because we wanted to explore a wide range of GCI types, we adopted a classification system for implicatures. For our experimental design, analysis, and presentation, we used Levinson's (2000) scheme, which classifies GCIs into three categories, as will be explained in detail below.

4.2.3.1 Q-based GCIs Levinson's class of Q-based implicatures consists of scalar implicatures, which arise from Grice's first Maxim of Quantity. This type of implicature is licensed

when a speaker uses a non-maximal value on some salient scale, thereby conveying that stronger values are either false or unknown. The Q-based implicatures we used were based on both entailment and non-entailment scales ("Horn scales" and "Hirschberg scales", respectively), and included rankings, quantifiers and modals, cardinals, and gradable adjectives. Examples of each of our four types of Q-based implicatures are given in (4).

(4) *Q-based implicatures*

a. Rankings

Irene: Who can register for the advanced seminar?

Sam: Juniors can register.

FACT: Anyone who has completed his or her first year of study can register.

b. Quantifiers and modals

Irene: How much cake did Gus eat at his sister's birthday party?

Sam: He ate most of the cake.

FACT: By himself, Gus ate his sister's entire birthday cake.

c. Cardinals

Irene: How many children does Lisa have?

Sam: Lisa has three children.

FACT: Lisa has quadruplets.

d. Gradable adjectives

Irene: How attractive is Kate?

Sam: She's pretty.

FACT: Kate was voted "World's Most Beautiful Woman" this year.

4.2.3.2 I-based GCIs Levinson's category of I-based implicatures arises from Grice's second Maxim of Quantity. In this category of GCIs, the implicature is based on the assumption that the speaker has said only what is necessary, leaving the hearer to infer a more specific and informative interpretation. For example, the most natural interpretation of the underlined portion of Sam's statement in (5a) below is the more specific interpretation that Rachel used the hammer to break the statue. We used a total of four different I-based implicatures including argument saturation, bridging inferences, co-activities, and conjunction buttressing. Examples of each are provided in (5).

(5) *I-based implicatures*

a. Argument saturation

Irene: I heard something big happened in the art studio yesterday.

Sam: Yeah! In a fit of rage, Rachel picked up a hammer and broke a statue.

FACT: After grabbing a hammer, Rachel angrily kicked a statue, causing it to fall over and break.

b. Bridging inferences

Irene: What happened when Sue came over?

Sam: She walked into the bathroom. The window was open.

FACT: The open windows are in the kitchen, and there are no windows in the bathroom.

c. Co-activities

Irene: Can the guys come to the reception?

Sam: No. George and Steve play squash at the gym until 6:00 everyday.

FACT: George plays squash at the YMCA until 6:00 daily, and Steve plays squash at SPAC until 6:00 everyday.

d. Conjunction buttressing

Irene: I understand that George has had a really rough year.

Sam: Yeah. Last month, he lost his job and started drinking.

FACT: George started drinking on the 15th of last month and lost his job on the 20th of last month.

4.2.3.3 M-based GCIs The last category of GCIs used in our study was Levinson's M-based implicatures. These implicatures arise from the form of the expression itself. M-based implicatures arise when a speaker chooses a marked way of describing a certain situation, thereby licensing the implicature that the standard description and, thus, the normal situation do not hold. From the set of M-based GCIs, we selected examples which we refer to here as verbal periphrasis and repeated verb and noun conjuncts. Examples of each are in (6) below.

(6): *M-based implicatures*

a. Verbal periphrasis
Irene: Why is Stephen so upset?
Sam: He caused Bill to die.
FACT: Stephen intentionally murdered Bill.

b. Repeated verb conjuncts
Irene: What happened at Doctor Witherspoon's office?
Sam: Sasha waited and waited for her appointment.
FACT: Sasha waited 5 minutes for her appointment at Doctor Witherspoon's office.

c. Repeated noun conjuncts
Irene: What did Joseph do after finishing the marathon?
Sam: He drank bottles and bottles of water.
FACT: Joseph drank one 20 oz bottle and one 16 oz bottle of water after finishing the marathon.

4.3 Experimental Tasks and Instructions There were three tasks associated with each stimulus item. Participants first read the short conversation between Irene and Sam and the FACT that followed, all presented on one slide. After reading this material, participants were prompted to determine whether the underlined portion of Sam's response was true or false, given the FACT. Lastly, they were asked to rate their confidence in their choice using a 4-point Likert scale.[4]

We had three instruction conditions that differed in two ways: (a) the training that the participants received before the experiment and (b) the way participants were asked to evaluate the truth of Sam's utterance. The first condition used a fictional character named Literal Lucy, a literal-minded individual, and participants were asked to evaluate the truth of the underlined portion of Sam's utterance from Literal Lucy's perspective. In the other two conditions, the training did not include Literal Lucy, and participants judged whether the relevant portion of Sam's utterance "interpreted literally is true or false" or "is true or false". Participants were reminded of how they were to interpret Sam's utterance on each stimulus slide after each conversation (see Sections 4.3.1 through 4.3.3 for illustration). We call the three conditions Literal Lucy, Literal No Lucy, and No Literal No Lucy, respectively. These three conditions will be explained in the sections below.

4.3.1 Literal Lucy Prior to reading the description of the experimental task, participants in the Literal Lucy condition were trained to interpret utterances from the perspective of Literal Lucy. By using this character, we hoped to draw upon participants' folk notion of interpreting literally, thereby suspending their natural adherence to principles of conversational cooperation. We hypothesize that the inclusion of this character would help to establish a shared, external perspective from which participants could evaluate the truth of the target sentences. Literal Lucy was described as someone who interprets language literally and, as a result, misinterprets things such as figurative language and indirect speech acts, illustrated in examples (7) and (8), respectively.

(7) Example of Literal Lucy's misinterpretation of figurative language

Frank: Brian just had a birthday, and I didn't realize how old he was.
Lucy: Really? How old is he?

[4] The analysis of the Likert data is not discussed in this paper.

Frank: He just turned 40, so now he's over the hill.
Lucy: Hill? Which hill? And when did he go over it?

(8) Example of Literal Lucy's misinterpretation of a speech act

Frank: Hey, Lucy, can you tell me what time it is?
Lucy: Yes, I can!
Frank: So...?

In (7), instead of interpreting the idiom *be over the hill* idiomatically (i.e. 'being past one's prime'), Literal Lucy interprets it literally (i.e. 'being over some particular hill'). In (8), Literal Lucy interprets Frank's question as literally asking about her physical abilities and not as an indirect request for the time. After demonstrating how Literal Lucy interprets examples such as (7) and (8), we instructed participants to adopt her way of interpreting utterances literally when evaluating the truth of statements in the experimental tasks. Finally, participants were presented with the example in (9) and told that Literal Lucy knows the FACT but that Irene and Sam may or may not know it, and that the possible difference between Literal Lucy's world knowledge and that of Irene and Sam does not matter for our purposes.

(9) Example of a conversation provided in the instructions

Irene: Hey, Sam. Do you know who wrote *Pride and Prejudice*?
Sam: <u>A British woman wrote it, and her last name was Austen.</u>
FACT: Jane Austen, a British woman, wrote *Pride and Prejudice*.

Here, Irene asks for the name of an author, and Sam describes the person who he thinks wrote *Pride and Prejudice*. Given that the FACT, which Literal Lucy knows, is truth-conditionally consistent with Sam's utterance, she would evaluate his underlined response as true. During the experiment, participants were instructed to assume the perspective of Lucy and to use the FACT as part of their basis for evaluating the truth of Sam's utterance. On each of the stimulus slides, when participants were prompted to evaluate Sam's utterance, they read the following instructions, "Given this FACT, Literal Lucy would say that the underlined sentence is: T/F", where T is true and F is false.

4.3.2 Literal No Lucy The Literal No Lucy condition differed from the Literal Lucy condition in that participants did not receive the Literal Lucy training prior to the introduction to the instructions for the experimental task. In this condition, the instructions began with the example in (9), this time used without reference to Literal Lucy. As in the Literal Lucy condition, participants were told that they should assume that the FACT was true and to ignore whether the speakers also knew the FACT. Since this condition did not make use of the Literal Lucy character, participants were asked to interpret the underlined sentence "literally". The instructions that occurred on the stimulus slides reflected this difference. Participants read the following instructions on each slide, "Given this FACT, the underlined sentence interpreted literally is T/F", where T is true and F is false.

4.3.3 No Literal No Lucy As in the Literal No Lucy condition, participants in this condition did not receive the Literal Lucy training prior to the introduction of the instructions for the experimental task. The only two differences between this condition and the Literal No Lucy condition are, first, that we omitted any use of the word "literally" in the instructions and, second, that we changed the wording of the task to "Given the FACT, the underlined sentence is T/F", where T is true and F is false.

4.4 Block Design and Presentation The materials were divided into 4 blocks, each block containing a single example of each of the 11 types of experimental GCIs and each of the 4 types of filler NCEs, and 7 control items (either 3 entailments and 4 contradictions or vice versa) for a total of 22 stimuli in each block. The ordering of the conversations was randomized, and the presentation of blocks was counterbalanced. Participants were tested individually using a self-paced reading paradigm. The tasks were performed on a computer in a sound-attenuated booth.

4.5 Participants A total of 74 native speakers of North American English from the Northwestern University community participated for pay or partial course credit. Two participants were excluded from the analysis, as they diverged by two standard deviations from the means on the control items (contradictions and entailments). Thus, data from 72 participants (24 per condition) were analyzed.

4.6 Interpreting the Data For the purpose of discussion, we present the data in terms

of percentage of false responses. False responses suggest that there is a truth-conditional incompatibility between the FACT and the underlined statement, therefore indicating that Sam's utterance was interpreted as false in light of the FACT presented. So, for example, we would expect a low number of false responses for entailments as in (2) above because the FACT entails the truth of Sam's statement and, therefore, is truth-conditionally consistent with it. For the GCIs, a high number of false responses would indicate that the GCI had been incorporated into the truth-conditional meaning of Sam's utterance. A low number of false responses for a GCI would indicate that the implicature had not been incorporated into the truth-conditional meaning of Sam's utterance.

For example, the FACT in (10), which includes the noun phrase *the entire cake*, is truth-conditionally inconsistent with the inference *Sam didn't eat all of the cake*, which is licensed by Sam's use of the quantified phrase *most of it* (where *it = the cake*) and is classified as a Q-based GCI.

(10) *Example of a scalar implicature*
Irene: How much of the cake did Gus eat at his sister's birthday party?
Sam: <u>He ate most of it.</u>
FACT: By himself, Gus ate his sister's entire birthday cake.

5. Results

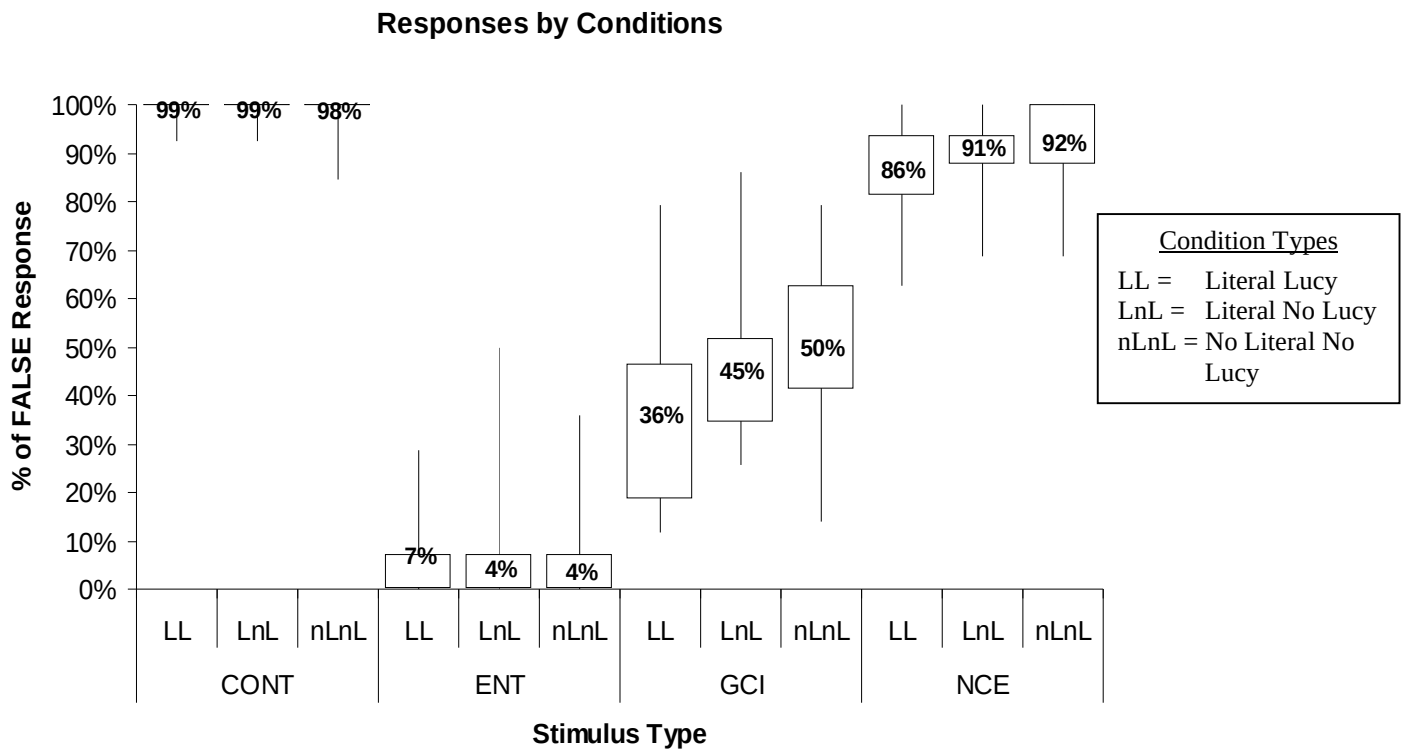
Recall the three questions we set out to answer in Section 3. First, are GCIs treated differently from other types of meaning? If GCIs pattern like entailments or contradictions, we can say with certainty that they are being uniformly incorporated into the truth-conditional meaning. If they emerge as a distinct class, this suggests that speakers do isolate a level of meaning that is distinct from WHAT IS SAID. Second, are all types of GCIs treated the same way? The answer to this question will help to determine whether all GCIs behave uniformly with respect to their incorporation into truth-conditional meaning. Third, do the various instructions utilized in the experiment enable us to determine what guides participants in the truth-evaluation task? The comparison of the three instruction conditions will reveal whether use of Literal Lucy's external perspective or a directive to "interpret literally" affects participants' ability to isolate the truth-conditional meaning of utterances.

5.1 Overall Pattern for All Stimulus Types

Using a logistic regression with GCIs as the initial baseline, we compared GCIs to each of the other stimulus types and found a significant effect of stimulus type. The average number of false responses for GCIs (44%) was significantly different from entailments (5%, $z = 4.98$, $p < .001$), contradictions (99%, $z = -8.37$, $p < .001$), and NCEs (93%, $z = -8.18$, $p < .001$).

Response patterns in the three different instruction conditions are shown in Figure (1). Entailments received the lowest median percentages of false responses (4% – 7%) and contradictions the highest (98% – 99%), showing that participants were in fact attending to and understanding the task. NCEs patterned similarly to contradictions (86% – 92%) but not as closely as one might have expected. In fact, it would appear that NCEs comprise their own, distinct class of meaning. Median false responses to GCIs ranged from 36% to 50%.

Figure 1: Boxplot of false responses per stimulus type for all instruction conditions



CONT = contradictions, ENT = Entailment

We also compared the three instruction conditions to one another and found a significant main effect of instruction condition. The average percentage of false responses for all stimuli in the Literal Lucy condition (57%) significantly differed from the Literal No Lucy condition (59%, $z = -2.09$, $p < .05$) and the No Literal No Lucy condition (60%, $z = -3.14$, $p < .001$). However, the Literal No Lucy and the No Literal No Lucy conditions did not significantly differ from one another ($z = 1.06$, $p = .29$), showing that the addition of an external perspective has a significant effect on the response pattern, but interpreting literally itself does not.

There was also a significant interaction between instruction type and stimulus type. Comparing contradictions and GCIs, for example, we found that instruction type seemed to have no effect on the response pattern in the contradiction items. In contrast with the GCIs, in which instruction condition did affect false response pattern, the lack of the effect of instruction type on contradictions is predicted, since the effect on truth-conditional meaning is independent of any contextual factors. Accordingly, participants should consider Sam's utterance as false whether they are interpreting his utterance literally or not, and whether from their own perspective or not. Entailments, on the other hand, were treated in the Literal Lucy Condition significantly differently than in the No Literal No Lucy ($z = 3.5$, $p < 0.001$) and the Literal No Lucy conditions ($z = 2.4$, $p < 0.05$). Participants in the different conditions may have treated entailments differently because participants in the Literal Lucy condition were more conservative in their acceptance of the truth of the utterances. For example, more Literal Lucy condition participants considered Sam's utterance in (11) to be false than participants in the other two conditions. One explanation for this difference is that they were differentiating between *university* (i.e. an institution of higher education that provides undergraduate and graduate education) and *college* (i.e. a division or school in a university or an institution that provides mostly undergraduate education).

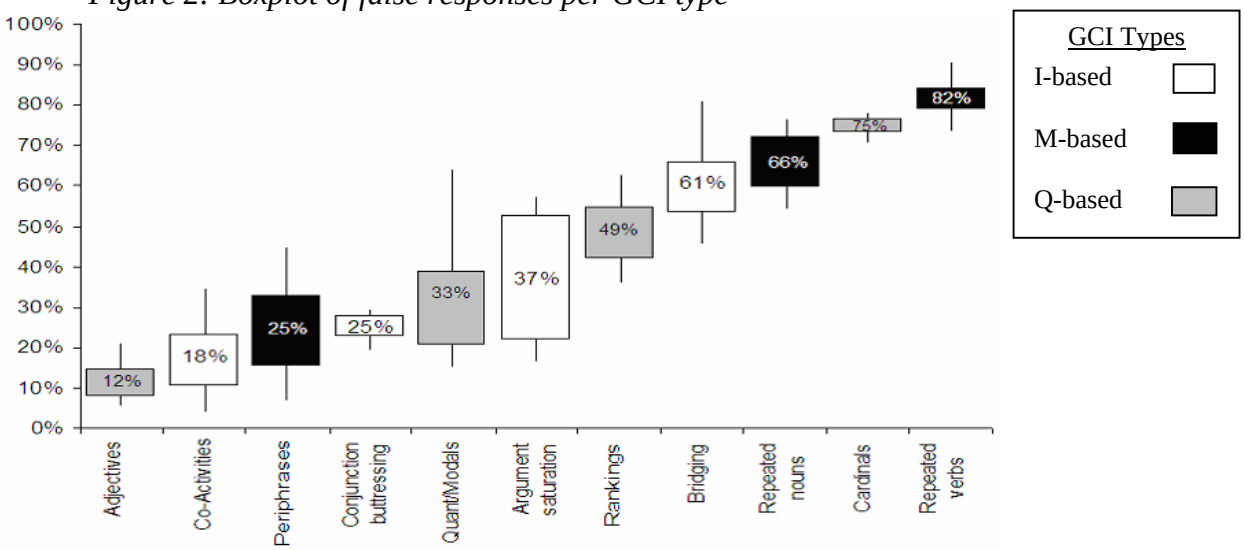
(11) *Example of an entailment with a relatively high percentage of false responses*
Irene: Why are John's parents so proud?
Sam: <u>He just got into college.</u>

FACT: John just got into Capitol University.

While participants in each condition rated contradictions similarly, participants in the Literal Lucy condition were significantly less likely to evaluate the GCIs in Sam's utterance as false than those in the other conditions ($z = -3.53$, $p < .0001$). This suggests that participants in the Literal Lucy condition were less likely than participants in other conditions to incorporate the GCIs licensed by Sam's utterance into the truth-conditional meaning of his utterance.

5.2 Results for GCIs Only Figure 2 below shows the median false responses for each of the GCI types across all conditions. Each of the GCI types is further marked for its classification according to Levinson's three implicature types. Figure 2 reveals two important points. First, different GCI types were treated differently by participants. The range of median false responses (12% – 82%) suggests that different GCI types differ in the degree to which they are incorporated by participants into truth-conditional meaning. Second, the difference in the way participants treated GCIs does not correspond to Levinson's classification system, in that they did not tend to cluster into I, Q, and M groups.

Figure 2: Boxplot of false responses per GCI type



The differences among Q-based implicatures as well as the M-based implicatures illustrate disparities within each GCI category in Levinson's system. For example, Q-based GCIs (gradable adjectives, quantifiers/modals, rankings, and cardinals) appear across the continuum, including the set with the lowest median percentage of false responses (gradable adjectives, 12%) and a set with one of the highest median percentages of false responses (cardinals, 75%). That is, some Q-based GCIs, e.g. cardinals, are often incorporated into the truth-conditional meaning, while other Q-based GCIs, e.g. gradable adjectives, are rarely incorporated into the truth-conditional meaning.

Likewise, M-based GCIs seem to form two groups based on the frequency with which they are incorporated into truth-conditional meaning. Repeated verb and noun conjuncts, for example, impact truth-conditions for many participants, while verbal periphrases do not.

6. Discussion and Conclusion

In this paper, we report on a new experimental paradigm that examined whether speakers reliably distinguish between Grice's notions of generalized conversational implicature (GCI) and WHAT IS SAID. Previous studies suggest that they do not, but none of them directly addresses truth-conditional meaning. The paradigm used in the current study included explicit instructions that guided participants in how to interpret the stimuli. The Literal Lucy condition had the greatest impact on participants' assessments of truth-conditions, and the effect was seen most in the GCI stimuli. More specifically, participants in the Literal Lucy condition were significantly less likely to judge our GCI stimuli to be false than participants

in the other two conditions and, thus, less willing to incorporate GCIs into truth-conditional meaning.

These results show that interpreting utterances from the perspective of a literal-minded third person enhances participants' ability to isolate a level of meaning corresponding to truth-conditional meaning. One explanation for these findings is that the adoption of an external perspective alone drives the effect and that the word "literal" plays no part. An alternative—and perhaps likelier—explanation is that assuming the perspective of a literal-minded individual is essential for interpreting literally, presumably through the suspension of cooperative conversational principles.

Past experiments that investigated the difference between WHAT IS SAID and WHAT IS IMPLICATED have included tasks that employed technical vocabulary, such as asking participants to determine what the speaker "meant"/"communicated" or even "what the sentence said". In these earlier studies, it is not clear what guided participants' judgments when they performed these tasks. Instead of directly addressing the question of WHAT IS SAID, we asked participants to focus on the truth of an utterance. Thus, by relying on their folk notion of interpreting literally, we avoided possible equivocation between the technical and non-technical senses of certain terms.

7. Future directions

One of the more surprising results of our study comes from the stimuli referred to as "necessary contextual elements" of meaning (NCEs), i.e. those contextually-determined elements of meaning that must be resolved for the establishment of truth conditions. In our study, we included four types of NCEs: deictics, ellipses, indexicals, and pronoun resolutions. Given the obligatory relationship between NCEs and truth conditions, we had initially predicted that they would pattern like other "purely semantic" elements of meaning, e.g. our contradiction control items. That is, if the FACT in a stimulus item presented information that is inconsistent with the interpretation of the NCE forced by the context, participants would judge it to be false, just as with those stimuli containing clearly contradictory information on an indisputably semantic level. However, this is not quite what we found. While NCEs received a very high rate of false responses (86% – 92%), they did receive slightly fewer false responses than the contradiction control items, which were at ceiling (98% – 99%). The difference between contradictions and NCEs was significant, suggesting that some participants, while making judgments in a literal mode of interpretation, are distinguishing even those contextually-supplied elements of WHAT IS SAID from those that are supplied strictly semantically. We intend to follow up on this finding in future work.

Our use of a wide variety of GCI types suggests several promising new avenues for research. For instance, we discovered that GCIs form a continuum with respect to the degree they were incorporated into truth-conditional meaning, a finding not predicted by either neo-Gricean or post-Gricean approaches. We are currently exploring possible contextual or grammatical factors that influence the degree of incorporation or the strength of an implicature. For example, in the case of scalar implicatures, the salience of alternatives or the nature of the scale itself may affect speakers' incorporation of scalar implicatures into truth-conditional meaning (Baker *et al*. 2008). The gradience of data from this experiment reveals that a more nuanced account of the relationship between GCIs and truth conditions is an important next step in the study of conversational implicature.

REFERENCES

Baker, Rachel E., Matthew Berends, Ryan Doran, Alex Djalali, Meredith Larson, Yaron McNabb, and Gregory Ward. 2008. An empirical investigation of the effects of scale type on the salience and the interpretation of scalar implicatures. Paper presented at the 2008 LSA Annual Meeting, Chicago, IL.

Bezuidenhout, Anne L. and J. Cooper Cutting. 2002. Literal meaning, minimal propositions and pragmatic processing. *Journal of Pragmatics* 34:433-456.

Bezuidenhout, Anne L. and Robin K. Morris. 2004. Implicature, relevance, and default pragmatic

inference. In Dan Sperber and Ira Noveck (eds.), *Experimental Pragmatics*. Hampshire: Palgrave MacMillan, 257-282.
Carston, Robyn. 2002. *Thoughts and Utterances: The Pragmatics of Explicit Communication*. Oxford: Blackwell.
Gibbs, Raymond and Jessica Moise. 1997. Pragmatics and understanding what is said. *Cognition* 62:51-74.
Grice, H. Paul. 1967. Logic and conversation. In Grice 1989: 22-40.
Grice, H. Paul. 1987. Further notes on logic and conversation. In Grice 1989: 40-58.
Grice, H. Paul. 1989. *Studies in the Way of Words.* Cambridge MA: Harvard University Press.
Horn, Laurence R. 1984. Toward a new taxonomy for pragmatic inference: Q- and R-based implicature. In Deborah Schiffrin (ed.) *Meaning, Form, and Use in Context.* Georgetown University Press, Washington D.C., 11-42.
Horn, Laurence R. 1993. Economy and redundancy in a dualistic model of natural language. *SKY 1993: 1993 Yearbook of the Linguistic Association of Finland*, 33-72.
Levinson, Stephen. 2000. *Presumptive Meanings: The Theory of Generalized Conversational Implicature*. Cambridge: MIT Press.
Nicolle, Steve and Billy Clark. 1999. Experimental pragmatics and what is said: A response to Gibbs and Moise. *Cognition* 69: 357-354.
Récanati, François. 1997. *Direct Reference: From Language to Thought*. Oxford: Blackwell.
Sperber, Dan and Deirdre Wilson. 1986. *Relevance: Communication and Cognition.* Oxford: Blackwell.