

Scalar Implicature in the Antecedent of Conditionals

Karin R. Howe

Department of Philosophy
Carnegie Mellon University
January 3, 2011

M.S. Thesis in Logic, Computation and Methodology
Under the supervision of
Dr. Mandy Simons
Dr. David Danks

Abstract

Embedded scalar implicatures have been the topic of much of the work on implicature since Grice's influential 1967 work on conversational implicature. Many people have seen issues surrounding these types of implicatures as reason to depart from the Gricean framework. Following Geurts & Pouscoulous (2009), we argue that this departure from the Gricean camp is premature, and that we should first try to gather empirical evidence for how people actually reason about utterances containing embedded scalar items before summarily dismissing Grice's theory as having nothing to say about these type of implicatures. For this thesis we conducted an experimental study modeled on Geurts & Pouscoulous' 2009 study of embedded scalar implicatures; however, our study focuses on scalar implicatures in the antecedent of conditionals, which Geurts & Pouscoulous do not examine. Using the survey responses to our experimental prompts, we answer the following two questions:

- a) Do scalar implicatures ever show up in the antecedent of conditionals?
- b) Do scalar implicatures always show up in the antecedent of conditionals?

In addition, we use the think-aloud portion of the experiment to gather insight into how people actually reason about these types of utterances, in order to generate an account of the reasoning patterns subjects employ in interpreting scalar terms in the antecedent of conditionals.

Table of Contents

Chapter I: Introduction	p. 1
Section I: Purpose of the Thesis	
Section II: Background	
Chapter II: Experimental Setup	p. 8
Section I: Pilot Study	
Section II: Final Experimental Design	
A. Experimental Prompts	
B. Instructions for Subjects	
Section II: Data Collection and Coding	
Chapter III: Analysis of Experiment	p. 17
Section I: Investigating Data From Survey Responses	
Section II: Investigating Data From the Think-Aloud Responses	
Section III: Inside the Black Box: Connecting the Think-Aloud Responses and the Survey Responses	
Section IV: Conclusions	
Chapter IV: Further Results: Behavior of Scalar Terms is Not Uniform	p. 34
Section I: <i>Warm</i> is Less Robust than <i>Some</i> or <i>Sometimes</i>	
Section II: <i>Sometimes</i> = <i>Occasionally</i> ?	
Section III: Further Research	
References	p. 49
Appendix I: Data Analysis	p. 51
Part I: Analysis of Coder Agreement	
Part II: Analysis of Survey Responses	
Part III: Analysis of Think-Aloud Responses	
Appendix II: Experimental Prompts	p. 69

Section I: Purpose of the Thesis

The purpose of this thesis is to investigate the behavior of scalar terms embedded within the scope of a conditional operator. Conditionals are an interesting case in terms of the literature on scalar implicatures, because there is no consensus on how they behave. Some accounts (e.g. Chierchia 2004) claim that scalar implicatures are suspended in the antecedent of conditionals, whereas other accounts (e.g. Levinson 2000) say that they are not suspended. Each of these accounts seems to be correct, in terms of the data that they each present. However, when you compare their data side by side, or if you tweak the sentences that they examine by changing the context or by changing the content of the consequent, then a less straightforward picture emerges, where the intuition is that scalar implicatures are suppressed in some conditionals, but not in others. Is there some way we can explain this phenomenon within a theoretical framework? Simons (20010) gives what she calls a "common-sense" explanation for the interpretation of conditionals with embedded scalar items. Briefly, the explanation is that the listener assigns an interpretation to the scalar item in the antecedent such that it improves the overall sensibleness of the utterance. We will be taking Simons' common-sense view as our starting point and backing it up with empirical results. Our hypothesis is that both local and global considerations come into play when "making sense" of the conditional, and that it is the connection between the antecedent and the consequent (and possibly also the context of the utterance) that produces pragmatic pressures which cause the scalar item to tip toward either a strengthened or unstrengthened reading.

We investigated the behavior of scalar terms in the antecedent of conditionals by conducting an empirical study, where we presented subjects with a variety of conditional statements containing the scalar terms *some*, *sometimes*, and *warm*. We collected two different kinds of data:

- 1) Survey data: Does the subject interpret the speaker as meaning the scalar term in the antecedent of the conditional in the strengthened sense (e.g., does the subject interpret the speaker to mean *some* as *some but not all*)?
- 2) Think-aloud data: How does the subject arrive at their interpretation of the speaker's utterance? What considerations does the subject consider to be relevant to making this interpretation?

In our study, we used a methodology adapted from Geurts & Pouscoulous (2009), where subjects were presented with a survey containing target utterances and potential implicatures, and were then asked to respond either "Yes," "No," or "I don't know" to the question of whether or not the speaker meant the potential implicature by his or her utterance. In addition, we added a think-aloud component to the experiment where we asked subjects to walk through the reasoning that they used in order to arrive at their answer to the survey prompt. Details of the experimental setup will be discussed in more detail in Chapter II. However, before we go on to that we will set the stage by briefly discussing some necessary background information on the theory of scalar implicature and the relevance of our work to that theory.

Section II: Background

The Gricean theory of conversational implicature has been highly influential in the field of pragmatics ever since Grice first introduced this theory in the William James lectures at Harvard University in 1967. In this lecture he developed a theory of conversational implicature based on one overarching principle, the Cooperative Principle, and four maxims falling under that principle: the maxims of Quantity, Quality, Relation and Manner. One important thread of the work on conversational implicatures that has followed Grice has focused on the phenomenon of scalar implicature, which most recently has been focused on issues with embedded scalar implicatures. Before discussing the issues that arise with embedded scalar implicatures, we first need to give a brief explanation of the theory of scalar implicature and how that theory fits into the Gricean framework. Scalar implicatures are based on informational scales, which involve two (or more) terms which are equally lexicalized, from the same word class, and from the same register, and which are then ordered from strongest to weakest; e.g. <all, most, some>, <always, often, sometimes>, or <hot, warm>. These informational scales were first discussed by Horn (1972), and thus are known as Horn scales. The standard formalization of Horn scales comes from Gazdar (1979):

If Q is a Horn scale of form $\langle \alpha_1, \alpha_2, \alpha_3, \dots, \alpha_n \rangle$, then each element in Q is logically weaker than any other element to its left in the following sense: given a pair of sentences S, S' identical except that S contains α_i , S' contains α_j and $i > j$, then, if α_i/α_j are not embedded under any operator S entails S' (as cited in Simons 2010).

As an illustration, consider the following set of sentences, which involve the Horn scale <all, most, some>:

- (1) The cat ate all of her food.
- (2) The cat ate most of her food.
- (3) The cat ate some of her food.

Here, (1) contains the scalar term *all*, which is to the left of the term *most* in the scale above. *Most* occurs in place of *all* in (2), and (1) entails (2). Likewise, (3) substitutes the term *some* for the terms *most* or *all*, both of which occur to the left of *some* in the scale above, and both (1) and (2) entail (3). Consequently, an utterance of either (2) or (3) implicates that (1) does not hold; that is, an utterance of either (2) or (3) implicates that the cat ate some (or most) of her food, but she did not eat all of her food. This implication is generated using Grice's Cooperative Principle and one of his conversational maxims, the first maxim of Quantity. These principles are given below:

Cooperative Principle

Make your conversational contribution such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged.

Chapter 1: Introduction

Quantity 1

Make your contribution as informative as is required (for the current purposes of the exchange) (Grice 1967, 1989).

Using the proposition expressed by a speaker's utterance as input, the Cooperative Principle and the first maxim of Quantity are used to generate a scalar implicature from the speaker's utterance. We will show how this reasoning works using statement (3) above as an example:

- (4) Linda is taking care of her cat Sylvia. The following exchange takes place between Linda and her mother Mary:

M: *Did the cat eat her food?*

L: *The cat ate some of her food.*

Linda's mother then reasons as follows: Linda said *The cat ate some of her food*, thereby expressing the proposition $p = \text{Linda's cat ate some of her [the cat's] food}$. Linda could have said *The cat has eaten all of her food*. Why did Linda say *some* rather than *all*? Since Linda's mother assumes that Linda is speaking cooperatively (and thus following the Cooperative Principle) and is also following the conversational maxims, she assumes that Linda is being maximally informative (following Quantity 1), and thus Linda is implicating:

- (5) *The cat has not eaten all of her food (i.e., The cat has eaten some but not all of her food).*

Grice's theory applies straightforwardly in unembedded scalar implicatures like (4). However, suppose Mary had then said to Linda's father Bob:

- (6) M: *Linda believes that the cat ate some of her food.*

When we try to apply the same analysis to (6) that we applied to (4), we run into problems. Intuitively, it seems plausible that Mary's utterance in (6) generates the following implicature:

- (7) *Linda believes that the cat ate some but not all of her food.*

However, suppose Linda's father tries to reason along the same lines as Linda's mother did in (4); i.e.: Mary said *Linda believes that the cat ate some of her food*, thereby expressing the proposition $p = \text{Linda believes that her [Linda's] cat ate some of her [the cat's] food}$. Mary could have said *Linda believes that the cat ate all of her food*. Assuming that Mary is speaking cooperatively and following the conversational maxims (and thus being maximally informative), Bob understands Mary to be implicating:

- (8) *Linda does not believe that the cat has eaten all of her food.*

However, this implicature is weaker than the implicature that seemed intuitively plausible in (7) above. (8) will be true whether or not Linda has any particular beliefs about whether cat has eaten all of her food or not, whereas (7) will only be true if Linda in fact does believe that the cat has eaten only some of her food.

Cases like this, where a purported implicature falls under the scope of a semantic operator (known as *embedded implicatures*), have been widely discussed in the literature as evidence against the Gricean theory of conversational implicature. According to Grice's theory, (6) should no longer generate the scalar implicature that was generated when the embedded clause uttered on its own in (4), because the proposition p that was expressed by Linda's utterance in (4) is not expressed by her mother's utterance in (6). According to Grice's theory, p serves as the input for the reasoning which generates a scalar implicature in (4), but p cannot serve as an input for the conversational reasoning in (6), since (6) does not express the proposition p . Thus, Grice's theory does not provide us with a mechanism to give an enriched interpretation to the embedded scalar term in (6).

This is problematic for Grice's theory, since as many authors have observed (Cohen 1971, Carston 1988, Levinson 2000, Chierchia 2004, Sauerland 2004, Geurts 2009, Simons 2010), apparent implicatures attaching to embedded clauses are not uncommon, and occur in a variety of different contexts, and under the scope of many different types of semantic operators. These issues have led many theorists (Levinson 2000, Chierchia 2004) to conclude that the Gricean framework should be abandoned, or at the very least heavily revised. One broad category of these theories is discussed by Geurts & Pouscoulous (2009), what they dub as the "conventionalist" view.¹ The idea behind this conventionalist view is that scalar implicatures aren't really inferences; rather, they are "special information packets which are associated with scalar expressions by linguistic convention" (Geurts & Pouscoulous 2009, p. 3). Thus, Geurts & Pouscoulous conclude that, in general,

[C]onventionalist theories tend to agree that, at least in upward-entailing environments, there is a preference for interpreting embedded *some*, in effect, as "some but not all," even if they disagree about how the construal and preference come about (Geurts & Pouscoulous 2009, p. 5).

Thus, scalar implicatures should occur freely and systematically in at least some embedded contexts, according to this theory.

Geurts & Pouscoulous (2009) tested this assumption through a series of experiments. The first experiment focused on scalar terms (namely, *some*) embedded under the scope of *think*, *want*, deontic *must*, and the universal quantifier *all*. Participants were presented with inference tasks like the following:

¹ For a list of the theorists which Geurts & Pouscoulous consider to be in this "conventionalist" group, see footnote 3 in Geurts & Pouscoulous (2009). Some of the key references are: Landman (1998, 2000), Levinson (2000), Recanati (2003), Chierchia (2004, 2006), and Fox (2007).

Chapter 1: Introduction

Emilie says:

*Betty thinks that Fred heard
some of the Verdi operas.*

Would you infer from this that Betty thinks
that Fred didn't hear all the Verdi operas?

yes no

In addition, participants were presented with one control item, in which *some* occurred in an unembedded context, as well as a number of filler items which had the same general format as the critical items, but which contained no scalar items. Geurts & Pouscoulous found that scalar implicatures (SIs) were endorsed over 90% of the time in the unembedded (control) items, but were endorsed at a considerably lower rate in all of the embedded cases (57.5% of the time in reference to *think*, and even lower in the other cases: *all*: 27%, *want*: 32%, *must*: 3%). These low rates of SIs obviously run counter to the conventionalists' view that scalar implicatures occur "systematically and freely" in embedded contexts, because if that was the case then there should be little or no difference between the embedded and unembedded contexts. Also interesting is the fact that the embedded cases were not endorsed at the same rate. As already noted, SIs were endorsed relatively frequently with *think*, very infrequently with *must*, and only about a third of the time with *all* and *want*. This is significant, because any conventionalist explanation as to the difference between the embedded and nonembedded cases runs up against difficulties in explaining the differences between the different embedded cases.

The importance of Geurts & Pouscoulous' experiment is twofold. First, and most obviously important, is the results themselves. Looking at these results, we can clearly see that scalar implicatures do not occur freely in embedded contexts. None of the embedded conditions achieved the robust results achieved in the unembedded condition, where scalar implicatures were endorsed over 90% of the time. Additionally, the claim that scalar implicatures occur "systematically" in embedded contexts can be seen to be false, since scalar implicatures were not endorsed at the same rate in all embedded contexts (compare the rates for *think* (57.5%) vs. *want* (32%) or *must* (3%) in the first experiment). Secondly, their experiments are important from a methodological point of view. Rather than simply using introspective evidence as most linguists do, Geurts & Pouscoulous examined what people really *do* when presented with these types of utterances. Empirical evidence seems like a very important piece of the puzzle, given that this is supposed to be a real-world phenomenon that we are attempting to explain. Thus, more experiments should be run to test the behavior of scalar items in other syntactically embedded contexts.

One such context is the conditional; specifically, scalar items embedded in the antecedent of a conditional. Conditionals have often been ignored in discussions of embedded scalar implicatures. Sauerland (2004) gives a Gricean treatment for a number of different types of embedded implicatures in his analysis, but ignores scalar terms embedded under the scope of a conditional operator. Likewise, Geurts & Pouscoulous (2009) tested the behavior of scalar items in a large variety of embedded contexts, but did not test them in the context of any conditionals. Levinson (2000) and Chierchia (2004)

Chapter 1: Introduction

are among the few who have had anything to say about scalar implicatures in the antecedent of conditionals. Levinson (2000) says, "The conditional provides some of the strongest apparent evidence for pragmatic intrusion" (p. 205). Pragmatic intrusion is the intrusion of pragmatically inferred content into the semantically driven truth-conditional content of the utterance (paraphrased from Huang, p. 189). Consider the following examples, given by Levinson:

- (9) If the USA won some of the Olympic medals, other countries must have got the rest.
- (10) If John owns three cars, then the fourth outside his house must belong to someone else.
- (11) If you reconnect the battery and turn the ignition key, the car will probably start.

In each of these cases, the statement is intuitively true when we give a strengthened reading to the antecedent, but is false otherwise. If we read the antecedent in (9) as *the USA won some and possibly all of the Olympic medals*, then it is unclear what "the rest" in the consequent refers to. However, if we read the antecedent as *the USA won some but not all of the Olympic medals*, then it all makes sense. Likewise, (10) makes more sense with a strengthened reading (*three and no more*) than it does with an unstrengthened reading (*at least three*). While (10) is not necessarily *false* given the unstrengthened reading (John could have more than three cars and the fourth car outside his house could still belong to someone else), it lacks the force (the fourth car *must* belong to someone else) that seems completely natural given the strengthened reading. Finally, (11) is simply false if the antecedent is not read as *If you reconnect the battery and then turn the ignition key*. If you tried to turn the ignition key without first reconnecting the battery, the car will not start. Thus, in all of these cases it seems that, as Levinson (2000) notes, "*if A then B* is intuitively true iff every way of verifying the entailments *and* the implicatures of A carries with it a verification of B" (p. 205).

Intuitively, it thus seems compelling on the basis of examples like these that you can get implicatures (scalar and otherwise) from clauses embedded in the antecedent of a conditional. Levinson even seems to be saying that implicatures in conditionals are relatively commonplace and unproblematic. However, as Simons (2009) notes, "[S]calar items in the antecedents of conditionals sometimes give rise to (local) scalar implicatures, and sometimes don't" (p. 22). To drive this point home, she compares one of Levinson's examples given above with another example which is structurally similar (has the same antecedent), but which leads to the opposite conclusion. Compare these two statements:

- (12) If John owns two cars, the third one in the driveway must be someone else's.
- (13) If John has two cars, he can lend us one for the afternoon.

As already observed, in (12) we read *two* as *exactly two*, or *two and no more*; however, (13) does not get the same strengthened reading, since there is no reason to believe that John's having more than two cars will change his ability to lend us one of them for the afternoon. Simons explains this phenomenon with what she calls the "common-sense"

Chapter 1: Introduction

view. The explanation in a nutshell is that the listener assigns an interpretation to the scalar item in the antecedent such that it improves the overall sensibleness of the utterance. Chierchia can also be seen to adopt a version of this common-sense account, according to Simons, in order to account for data that fails to accord with the predictions given by his account of scalar implicatures, which is that scalar implicatures will be suppressed in the antecedent of conditionals. Chierchia's response to (12) is that we make sense of this utterance by restricting our attention to "sets of worlds from which people with no more than two cars are excluded" (Chierchia, p. 67). This accommodation is added to avoid a "near contradiction," according to Chierchia.

Simons is inclined to accept the common-sense view at face value, as am I. Simons' argues as follows:

[T]he interpreter of a conditional recognizes the antecedent of the conditional as a linguistic unit with a specific function within the discourse. She seeks to assign to it an interpretation which maximizes the cooperativity of the speaker. But the interpretation is not necessarily linear. We need not assume that the interpreter assigns to the antecedent an interpretation to the antecedent entirely independently of the attempt to make sense of the utterance as a whole. Rather, it seems plausible that the interpreter assigns to the antecedent an interpretation which improves the overall sensibleness of the utterance (p. 24).

As we can see from the above quote, this assumption that a speaker's utterance "makes sense" lines up with Grice's assumption that the speaker is speaking cooperatively. This common-sense view can then be seen to fit with the data presented by both Levinson and Chierchia. However, unlike either Levinson's or Chierchia's views, this view will also predict that the behavior of scalar terms in the antecedent of conditionals should not be expected to behave in a uniform manner, either in terms of always being strengthened, or never strengthened. Thus, the hypothesis is that both local and global considerations come into play when "making sense" of the conditional, and that the contents of the conditionals (the content of the antecedent and the consequent, and the connections between these contents) produce pragmatic pressures in favor of different interpretations in different cases.

Section I: Pilot Study

Following Geurts & Pouscoulous (2009), we devised an experiment to test subjects' interpretations of conditional sentences whose antecedents contain a scalar item. The sentences were constructed using three different scalar terms (*some*, *sometimes*, and *warm*). Following their model we also included sentences containing unembedded scalar items (using the same scalar terms as in the target prompts) to serve as control prompts. A potential pool of 24 target prompts (8 for each scalar term) was prepared, with the plan to narrow this list down to a total of 15 target prompts. In addition, we devised 12 control prompts (4 for each scalar term). The 24 target prompts were each judged by us to fall into one of four categories. As we discussed in Chapter I, our hypothesis was that the contents of the conditionals (the content of the antecedent and the consequent, and the connections between these contents) would produce pragmatic pressure in favor of different interpretations in different cases. We wanted to present subjects with a range of conditionals with different pragmatic pressures in order to bring out the effects of these pragmatic pressures. Thus, in devising our potential pool of target prompts we created prompts that we judged to lean toward a strengthened or unstrengthened interpretation, as well as target prompts that we judged to be "neutral," in that they did not *a priori* seem to lean one way or the other. We also devised a group of "truth functionally" strengthened target prompts, prompts which *require* a strengthened interpretation in order to be coherent.

A pilot study was conducted in order to test the general effectiveness of the potential experimental prompts, and to help narrow down the list of target prompts to the desired number. Six pilot surveys were created and administered to six pilot subjects. Each pilot survey contained all 24 of the potential target prompts, presented in a different pseudo-random order² for each subject. In addition, each pilot survey contained a random choice of 3 out of the 12 potential control prompts (1 for each scalar term), and all 12 of the potential control prompts were presented to at least one pilot subject. The pilot subjects were thus each presented with a total of 27 prompts, which they were asked to respond to with either a "Yes," "No," or "I don't know" answer. A sample prompt is shown below:

Question 13

Sally and Jim have four cats. Sally says:

If some of the cats get a treat, then the others will be jealous.

Does Sally mean:

If some but not all of the cats get a treat, then the others will be jealous.

Yes No I don't know

² Pseudo-random ordering of prompts was generated by assigning all the prompts a number and then using a random sequence generator (<http://www.random.org/sequences/>) to generate a random ordering of the numbered prompts. This ordering was then adjusted if it seemed like too many similar prompts were placed next to each other, simply by chance.

In all the experimental prompts, "Yes" responses correspond to strengthened interpretations and "No" responses correspond to unstrengthened interpretations, where by strengthened interpretations we mean interpretations of a scalar term that exclude the higher alternatives on the scale. For example, in the prompt above, the proposed strengthened interpretation of *some* is *some but not all*, excluding the scalar alternative *all* from the possible meanings of *some*. In contrast, an unstrengthened interpretation of *some* in the prompt above would include the possibility that all the cats got a treat (e.g.; Does Sally mean: *If some and possibly all of the cats get a treat, then the others will be jealous*).

In addition to being asked to choose either a "Yes," "No," or "I don't know" answer, we also asked the pilot subjects to think aloud as they answered the survey questions in order for us to get a better picture of how they were reading and interpreting the prompts. The idea was that having the pilot subjects do this would help us identify any problematic features of the prompts, and would also help us determine whether or not subjects were generally reacting to the prompts in the way that we thought they would (and if not, would help us understand why not).

The results of the pilot study were very helpful and encouraging. In general, subjects reacted to the target prompts in line with the way that we categorized them (strengthened and "truth functionally" strengthened prompts generally receiving a "Yes" answer and unstrengthened prompts generally receiving a "No" answer, while the neutral prompts were more variable). Where subjects diverged from our expectations, the think-aloud recordings were helpful in clarifying why this was the case and how we should respond to it, either by clarifying or expanding on the context provided in the prompt, or perhaps by getting rid of the prompt altogether. However, we also discovered that the think-aloud portion of the pilot study gave us interesting and diverse data about the reasoning processes that the subjects used in order to arrive at their interpretations of the target prompts, data that would not be recoverable using the survey data alone. Two different subjects might arrive at the same interpretation using different considerations or reasoning processes. Thus, we decided to include a think-aloud protocol in the final experimental design in order to gather some qualitative data on the considerations or reasoning processes that subjects use in interpreting purported scalar implicatures (embedded or otherwise) in addition to the data generated by their survey responses.

Section II: Final Experimental Design

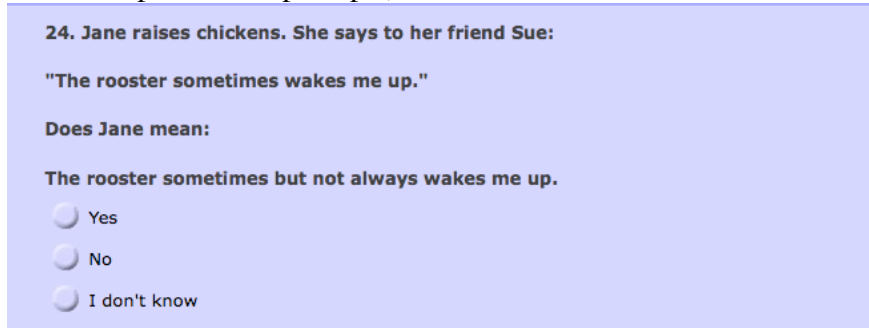
A: Experimental Prompts

In the final experimental design, subjects were presented with 39 prompts (15 target prompts, 12 control prompts, and 12 distractor prompts) presented in a pseudo-random order³ and asked to respond either "Yes," "No," or "I don't know" to the prompt. Each type of prompt will be discussed separately below.

³ All subjects were presented the same set of prompts in the same order.

Control Prompts

Unembedded scalar implicatures were used as control items. These were used to test if subjects were generating scalar implicatures at all, and served as a basis for comparison with the target items. There were 12 control items, 4 for each of the three scalar items (*some*, *sometimes*, and *warm*). A sample control item is given below (see Appendix II for a full list of all experimental prompts).



24. Jane raises chickens. She says to her friend Sue:
"The rooster sometimes wakes me up."
Does Jane mean:
The rooster sometimes but not always wakes me up.
 Yes
 No
 I don't know

Target Prompts

As described in Section I, the target prompts for this experiment can be separated into four different categories:

- **S:** prompts we judged to lean toward a strengthened interpretation (e.g., the scalar term *sometimes* is understood in the context of the prompt to mean *sometimes but not always*)
- **U:** prompts we judged to lean toward an unstrengthened interpretation (e.g., the scalar term *sometimes* is understood in the context of the prompt to mean *sometimes and possibly always*)
- **N:** "neutral" prompts; prompts that don't *a priori* seem to lean one way or the other
- **TF:** "truth functionally" strengthened prompts; prompts that *require* a strengthened interpretation in order to be coherent

Subjects were presented with 15 target prompts according to the following distribution: 2 strengthened prompts (1 each involving the scalar terms *sometimes* and *warm*), 4 unstrengthened prompts (1 each involving the scalar terms *sometimes* and *warm*, and 2 involving the scalar item *some*),⁴ 3 truth functionally strengthened prompts (1 each involving each of the scalar terms), and 6 neutral prompts (2 each involving each of the

⁴ One of the unstrengthened target prompts involving the scalar term *some* was originally intended as a strengthened target prompt, but was determined upon reflection to lean more towards an unstrengthened interpretation and so it was moved to that category. However, since data collection had already begun it was impossible to replace it with another strengthened prompt and thus we unavoidably ended up with 2 unstrengthened target prompts involving the scalar term *some*, and no strengthened target prompts of this type.

scalar terms). Examples of each category are given below, along with a brief discussion of why we judged them to fall into that category.⁵

Strengthened Prompt:

36. Jim lives in England with his wife and two daughters. He and his wife are planning a trip to visit his sister in Pennsylvania. Jim says:

"If we want it to be warm when we visit, we should go in April."

Does Jim mean:

If we want it to be warm but not hot when we visit, we should go in April.

Yes

No

I don't know

This prompt was judged to lean toward a strengthened reading because of a variety of contextual factors given in the prompt. Having Jim say that they should visit in April if they wanted the weather to be warm was intended to elicit a strengthened interpretation from the subjects, since April is a spring month, and is generally warm but not too hot. Giving Pennsylvania as the destination for their visit was meant to increase the likelihood of the subjects being able to arrive at this interpretation given their familiarity with the weather in Pittsburgh, PA in April.⁶ The choice of England (which has a fairly mild climate) as the point of departure was also deliberately chosen as opposed to some place with a colder perceived climate, such as Norway, which might caused the prompt to lean in the other direction, toward an unstrengthened interpretation.

Unstrengthened Prompt:

18. A professor and his TA are having a meeting. The professor says:

"If some of the students fail the exam, then I will give an extra credit assignment."

Does the professor mean:

If some but not all of his students fail the exam, then I will give an extra credit assignment.

Yes

No

I don't know

⁵ See Appendix II for a full list of all experimental prompts

⁶ This turned out to be overly optimistic, given the tendency for subjects to see Pittsburgh as a generally cold and miserable place (as is discussed later in Chapter IV, Section I), in contrast with the investigators, who had both spent considerable time in upstate New York and thus had perhaps a different view of the relative warmth of Pittsburgh compared to other places.

Chapter II: Experimental Setup

This prompt was judged to lean toward an unstrengthened interpretation, since it seemed unlikely that having more students (all the students) fail the exam would change the professor's mind about giving an extra credit assignment, although such an interpretation was certainly possible if subjects thought that the professor might do something else in that case (such as give a new exam).

Neutral Prompt:

29. The operating manual for a piece of machinery says:

"If the Z-plate is warm, turn the S-dial counter-clockwise."

Does the manual mean:

If the Z-plate is warm but not hot, turn the S-dial counter-clockwise.

Yes

No

I don't know

The neutral prompts are ones where we tried as much as possible to remove the pragmatic pressures that might push the prompt one way or the other, either toward a strengthened or an unstrengthened interpretation. This was mainly done either by giving as little background context as possible (without making it obvious that this is what we were doing) or by giving contexts that are less likely to be familiar to the subjects and part of their everyday life (e.g. contexts involving chicken farming). In addition to impoverishing the background context, we also tried to devise the neutral prompts in such a way that the connection between the antecedent and the consequent was less obvious. The example above employs both of these methods of making the prompt neutral. The background context is very impoverished, in that there is no clue what type of machine the instruction manual is referring to, or whether it is good or bad if the Z-plate is warm or not, and whether it would be better or worse if it got warmer than that. Also, the connection between the antecedent and the consequent is underspecified. Does turning the S-dial counter-clockwise make the Z-plate colder, or warmer? There's no way to tell. Thus we expected it to be particularly interesting to see what interpretations people give to these type of prompts, and what sorts of considerations and reasoning processes they use to arrive at these interpretations, and how this compares with how they respond to the other types of prompts.

Truth Functionally Strengthened Prompt:

21. A professor and her TA are having their weekly meeting. The professor says:

"If we give some of the students an extension, the others will be upset."

Does she mean:

If they give some but not all of the students an extension, the others will be upset.

- Yes
- No
- I don't know

The "truth functionally" strengthened prompts are prompts which require a strengthened interpretation in order to be coherent, as this prompt illustrates. There is really no other way to interpret this prompt. If *some* is not interpreted as *some but not all*, who are *the others*? If all of the students are given an extension, then there will be no "others" to be upset! The only way such an interpretation would make sense would be to see "the others" as an empty set, which can be done, but then one wonders what the point of referring to them is, and thus this interpretation seems fairly unlikely.

Distractor Prompts

The distractor prompts were formatted the same as the control and target prompts, and like the target prompts, contained an embedded scalar item (*some*). However, in the distractor prompts *some* was embedded under the scope of different propositional attitude predicates (e.g., *be surprised*, *regret*). A sample distractor prompt is given below:

3. Prof. Smith held a party at his home for students interested in majoring in linguistics. He was quite pleased with the turn-out, and reported:

"I was surprised that some of the students from my intro class showed up."

Did Prof. Smith mean that some but not all of the students from his class showed up?

- Yes
- No
- I don't know

Distractor prompts were included for the purpose of disguising the true aims of the experiment and to break up "learning effects," and thus were not included in the analysis of the experiment.

B: Instructions for Subjects

Subjects were presented with the experimental prompts one at a time and asked to "walk through" the reasoning process that they used in order to arrive at the answer ("Yes,"

"No," or "I don't know") that they chose for that prompt before proceeding to the next prompt. Audio recordings were made of each subject's think-aloud responses, which were then coded according to a coding manual (see Section III for a discussion of the coding process).

Section III: Data Collection and Coding

We collected survey data from 26 participants (10 male, 16 female) and audio recordings of the think-aloud portion of the experiment were collected for 24 of the participants (9 male, 15 female).⁷ The think-aloud data was then coded according to the following coding manual:

H	Subject references a higher item on the scale; i.e., <i>all</i> or <i>most</i> rather than <i>some</i> , <i>hot</i> rather than <i>warm</i> , or <i>always</i> rather than <i>sometimes</i> .
L	Subject references a lower item on the scale; i.e., <i>few</i> or <i>none</i> rather than <i>some</i> , <i>cold</i> rather than <i>warm</i> , or <i>never</i> rather than <i>sometimes</i> .
A	Subject references alternate utterances the speaker might have made instead of the utterance given in the prompt.
P	Subject considers possible alternate states of the world; i.e., subject references hypothetical alternate scenarios that might be (could have been) the case rather than the scenario given in the prompt.
CS	Subject explicitly exhibits sensitivity to context, either by making reference to general world knowledge that he or she considers to be relevant in interpreting the speaker's utterance or by embellishing the context beyond what is given in the prompt. Alternatively, the subject explicitly asserts that he or she lacks information about the context or the speaker that is necessary for making an interpretation of the speaker's utterance.

These codes were generated based on the data from the pilot study, and represent the most common features we found in the subjects' responses in the pilot study. These codes are not mutually exclusive. In other words, a response may be coded with as many codes as is appropriate given the content of the response. Thus, some responses may receive several codes, while other responses receive only one or two, or perhaps even none. Coder agreement was determined by comparing the independent coding results of the primary coder and an outside coder on a sample of the think-aloud recordings, and was found to be within acceptable parameters.⁸ We will illustrate how these codes were applied to the data using the following sample responses to the following prompt:

⁷ Data from the think-aloud portion of the experiment was lost for two of the subjects due to equipment malfunction during data collection.

⁸ For details of the analysis of coder agreement see Analysis 1.1 in Appendix I.

29. The operating manual for a piece of machinery says:
"If the Z-plate is warm, turn the S-dial counter-clockwise."
Does the manual mean:
If the Z-plate is warm but not hot, turn the S-dial counter-clockwise.

Yes
 No
 I don't know

1. Subject 101: Um, I'm going to say "no," because it [the original utterance] suggests that the Z-plate being warm is a strange condition, and it should be cool normally – so if it was hot you'd also want to [turn the S-dial counter-clockwise].
 - In this response the subject clearly references both the higher item on the scale ("hot") and a lower item on the scale ("cool"), and thus it was coded as both **H** and **L**. The subject also considers hypothetical alternate scenarios that might be the case rather than the scenario given in the prompt ("if it was hot you'd also want to...") and thus is coded as **P** as well. Finally, this response was also coded as **CS** because of the way the subject embellishes the prompt (there is nothing explicit in the prompt that indicates that the Z-plate being warm is a "strange condition").
2. Subject 110: Yes. So, they would have explicitly mentioned if it is hot, then do something else, so it's true – if the plate is warm enough, then turn the dial counter-clockwise. (H, A)
 - In this response the subject mentions just the higher item on the scale ("hot"), so this response receives an **H** code. Additionally, the subject references an alternate statement that might have been given in the manual ("they would have explicitly mentioned if it is hot, then do something else"), so this response is also coded as **A**.
3. Subject 118: Well, again this is one of those ones where there's no background information to really tell you. I mean ... if the Z-plate were a plate on a stove, um, then it's – what you'd be aiming for is to get it really hot, um, so in that case it might mean *if the Z-plate is warm but not hot*. On the other hand, ah, maybe the Z-plate is some kind of cooling device, and it's bad that it's warm – in which case it would be aiming to say *if the Z-plate is warm instead of cold, then turn the S-dial counter-clockwise* and that would solve the problem. Um so, yeah, without background information it's very hard to say, um, except that I guess I'd expect if we were talking about something that's supposed to be hot they'll say *if the Z-plate is not hot enough*, or is *too cool*, um, it would be unusual for them to say *warm*. I guess usually when I hear the word *warm* I think it means *warm* instead of *cold* rather than *warm but not hot*, so I'm going to go "no" with that one, but the lack of background information makes it hard to say.
 - This subject's response is extremely unusual, in that received all of the codes. Straightforwardly, it receives an **H** and an **L** code because the subject clearly references both the higher item on the scale ("hot") and a lower item on the scale ("cold"). The subject also clearly references an

Chapter II: Experimental Setup

alternate statement that might have been given in the manual ("if we were talking about something that's supposed to be hot they'll say *if the Z-plate is not hot enough*, or is *too cool*"), so this response is coded as **A**.

However, the subject also considers several different hypothetical scenarios that might be the case rather than the scenario given in the prompt ("if the Z-plate were a plate on a stove, um, then it's – what you'd be aiming for is to get it really hot," "maybe the Z-plate is some kind of cooling device, and it's bad that it's warm," "if we were talking about something that's supposed to be hot"), so the response is also coded as **P**. Finally, the response is coded as **CS** for two reasons: 1) the subject embellishes the context beyond what is given in the prompt ("if the Z-plate were a plate on a stove," "maybe the Z-plate is some kind of cooling device"), and 2) the subject explicitly states that she lacks relevant background information several times.

As you can see, subjects use a variety of considerations and reasoning processes in order to arrive at their interpretations for this prompt. We will discuss the insights into subjects' reasoning processes that can be gleaned from the think-aloud data in more detail in Sections II-IV of the next chapter. First, however, we turn in Section I to questions that can be answered using the survey data alone.

Section I: Investigating Data from Survey Responses

In analyzing the data from the survey responses,⁹ we began by asking the following two top-level questions:

1. Do scalar implicatures ever occur in the antecedent of conditionals?

2. Do they always occur in the antecedent of conditionals?

The distribution of the survey responses of the 26 participants to the target and control prompts is given in Table 1 below:¹⁰

Table 1	Yes	No	I don't know
Target	52.6%	42.3%	5.1%
Control	74.4%	22.4%	3.2%

Clearly, the answer to the first question is yes; scalar implicatures *do* occur in the antecedent of conditionals, since subjects responded "yes" to 52.6% of the target items, and a "yes" response to a survey question corresponds with a strengthened reading of the scalar term in the antecedent of the conditional given in the prompt. Likewise, it is equally clear that the answer to the second question is no; scalar implicatures do not always occur in the antecedent of conditionals, since subjects responded "no" or "I don't know" to 47.4% of the target items. Thus, our original intuitions are confirmed: scalar items in the antecedent of conditionals sometimes get a strengthened reading, and sometimes don't.

Interestingly, similar conclusions could be drawn for the unembedded control prompts, although to a lesser degree than in the target prompts. While almost 75% of the control prompts received a strengthened reading, there were a surprising number of "no" and "I don't know" responses to the control prompts. To get a clearer idea of why the control prompts received such a surprisingly low rate of strengthened interpretations, we examined the distribution of the survey responses to the target and control prompts, broken down by scalar term. The results are given in Table 2 below:¹¹

Table 2		Yes	No	I don't know
Target	<i>Some</i>	50.8%	45.4%	3.8%
	<i>Sometimes</i>	64.6%	33.1%	2.3%
	<i>Warm</i>	42.3%	48.5%	9.2%
Control	<i>Some</i>	96.2%	1.9%	1.9%
	<i>Sometimes</i>	91.3%	6.7%	1.9%
	<i>Warm</i>	35.6%	58.7%	5.8%

Looking at the results for the control and target prompts when broken down by scalar term, we see that in the control prompts containing *some* and *sometimes* there is a definite preference for a strengthened reading (over 90% of these prompts received a strengthened reading). The target prompts containing *some* and *sometimes* were strengthened at a

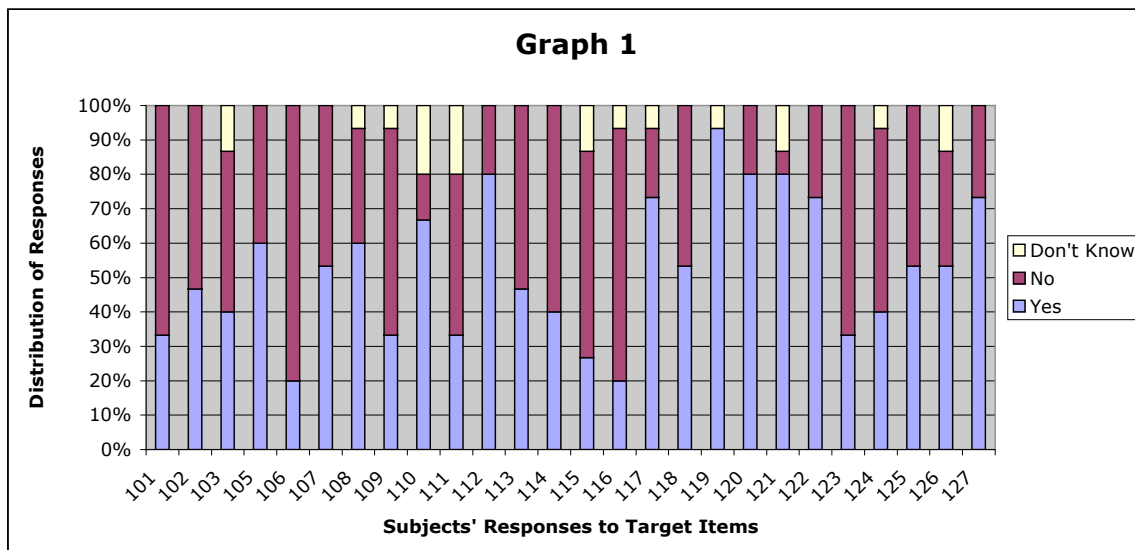
⁹ Data from distractor prompts was excluded from all analysis for this experiment.

¹⁰ For details of this analysis see Analysis 2.1 in Appendix I.

¹¹ For details of this analysis see Analysis 2.2 in Appendix I.

significantly lower rate than the corresponding control prompts ($p < .001$),¹² but were still strengthened in over half of the items. However, *warm* doesn't systematically give rise to scalar implicatures even - perhaps especially! - in the unembedded control prompts. This suggests that purported scalar terms may differ significantly in their behavior. We will look at this question in more depth in Chapter IV.

One possible concern with the results for the target prompts (that they sometimes get a strengthened reading, and other times do not) is that perhaps this result is a consequence of two different modes of reading scalar terms in the antecedent of conditionals; one mode which is to always strengthen, and another mode which is to never strengthen. In other words, the worry is that the answer to the second question is both yes and no; yes, in that some people always strengthen scalar terms in the antecedent of conditionals, and no, in that other people never strengthen scalar terms in the antecedent of conditionals, and this difference gets averaged out when we do the cross-subjects analysis. However, a within-subjects analysis of the distribution of the survey responses to the target prompts shows that this worry is clearly misplaced, as Graph 1 shows:¹³



All of the subjects gave strengthened readings to some of the target items, and only one subject¹⁴ showed a tendency to almost always give a strengthened reading. The mean number of "yes" responses was 7.6, and the mode was 5. Additionally, there were no statistically significant gender differences¹⁵ in the responses of men versus women to the target items.¹⁶ Thus, we see that the variability in subjects' responses to the target items is

¹² For details of this analysis see Analysis 2.3 in Appendix I.

¹³ For details of this analysis see Analysis 2.4 in Appendix I.

¹⁴ Subject 119: yes=14, don't know=1. See Analysis 2.4 in Appendix I.

¹⁵ For details of this analysis see Analysis 2.5 in Appendix I.

¹⁶ There was a slight gender effect in the response rates to the control items ($p = .033$), which can be explained entirely by the difference in the number of "I don't know" responses in women versus men (women were more likely than men to say "I don't

in fact as we hypothesized: people strengthen scalar items in the antecedent of some conditionals, but not in others.

Our hypothesis was that the contents of the conditionals (the content of the antecedent and the consequent, and the connections between these contents) produce pragmatic pressure in favor of different interpretations in different cases. Recall that in designing the experiment, we wanted to present subjects with a range of conditionals with different pragmatic pressures in order to bring out these effects. We grouped the target prompts into the four different categories listed below:

- **S**: prompts we judged to lean toward a strengthened interpretation
- **U**: prompts we judged to lean toward an unstrengthened interpretation
- **N**: "neutral" prompts; prompts that don't *a priori* seem to lean one way or the other
- **TF**: "truth functionally" strengthened prompts; prompts that *require* a strengthened interpretation in order to be coherent

The distribution of yes/no/don't know answers for the 15 target prompts, broken down by the four categories given above, is reported in Table 3 below:

Table 3	Yes	No	I don't know
S	73.1%	21.2%	5.8%
U	33.7%	63.5%	2.9%
N	37.8%	53.8%	8.3%
TF	93.6%	5.1%	1.3%

As we can see from these results, the prompts in the different categories (except for the unstrengthened vs. neutral) behave in significantly different ways.¹⁷ In general, subjects responded to the target prompts in the directions that we hypothesized that they would; subjects showed a systematic tendency to strengthen the target prompts that we categorized as leaning toward a strengthened reading (with an even more marked tendency to do so in the case of the "truth functionally" strengthened prompts), while showing a tendency toward an unstrengthened reading in the target prompts that we categorized as leaning toward a unstrengthened reading. Interestingly, the "neutral" prompts appear at a first glance to behave in much same way as the prompts that we judged to lean toward an unstrengthened reading. However, a comparison of the six neutral prompts also shows a statistically significant difference in the frequency of Yes vs. No responses in the six items ($p < .05$), unlike the same comparison run on the items in the other three categories (**S**: $p = .052$; **TF**: $p = .744$; **U**: $p = .087$).¹⁸ The fact that the behavior of the neutral items is not uniform (unlike the behavior of the items in the other three categories) suggests that the "neutral" items are not all equally neutral, and in fact

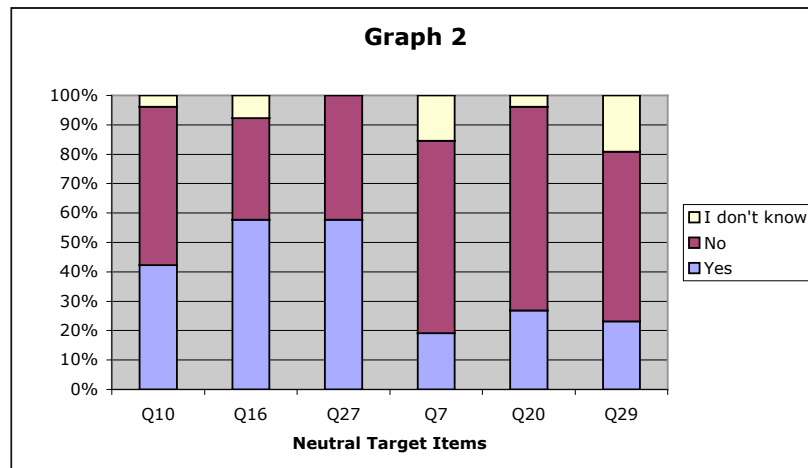
know"). See Analysis 2.6 in Appendix I for the distribution of responses of males vs. females to the control items.

¹⁷ A pairwise comparison of the four different categories showed that, with the exception of **U** vs. **N** ($p = .296$), the difference in the frequency of "Yes" vs. "No" responses in each of the categories was statistically significant ($p < .001$). For details of this analysis see Analysis 2.7 in Appendix I.

¹⁸ For details of this analysis see Analysis 2.8 in Appendix I.

Chapter III: Analysis of Experiment

we can see that this is the case by looking at the distribution of the responses to the different neutral target items, shown in Graph 2 below:¹⁹



Some questions do in fact get the ~50/50 response that one might reasonably expect in the neutral prompts, but the rest lean toward an unstrengthened reading, which helps explain some of the similarity between the neutral and unstrengthened categories. Specifically, questions 10, 16, and 27 have an approximately 50/50 distribution of Yes/No responses, while questions 7, 20, and 29 lean toward an unstrengthened reading.²⁰

The difference between the different neutral prompts is best explained by looking at the prompts themselves. The six neutral prompts are given below:²¹

- Q10: "If we sometimes have lambs that need to be bottle fed, we will post it on our website."
- Q16: "If foxes kill some of the chickens, then we will have less eggs to sell."
- Q27: "If you sometimes have no time to cook, then our frozen dinners are a healthy alternative."
- Q7: "If you want your pie to be warm, please let me know."
- Q20: "If some of the chickens are black, then there will be an announcement."
- Q29: "If the Z-plate is warm, turn the S-dial counter-clockwise."

The first thing to note is that all of the neutral prompts containing the scalar term *warm* ended up in the group that acted more like the unstrengthened prompts. This is not

¹⁹ For details of this analysis see Analysis 2.9 in Appendix I.

²⁰ A Chi Square analysis run on 10, 16, and 27 shows no statistically significant differences between the distribution of Yes/No answers to these items ($p = .399$). Likewise, there is no statistically significant difference between 7, 20, and 29 ($p = .899$). Additionally, a Chi Square analysis run on the items originally judged to lean toward an unstrengthened reading along with the "neutral" items 7, 20, and 29 shows no statistically significant difference between any of these items ($p = .211$), thus justifying classifying 7, 20, and 29 as also being more unstrengthened-leaning than "neutral." For details of this analysis see Analysis 2.10 in Appendix I.

²¹ For the full prompts see Appendix II.

particularly surprising, given the trend we have already observed for subjects to be more likely to interpret prompts containing *warm* as meaning *warm and possibly hot* (or *warm but not cold*), rather than *warm but not hot*. Thus it is not unexpected that the neutral prompts containing the scalar term *warm* would lean more in an unstrengthened direction. Secondly, we note that all of the neutral prompts containing the scalar term *sometimes* ended up in the ~50/50 group. In analyzing this result, a closer examination of the target prompts containing the scalar term *sometimes* can help explain the behavior of these two neutral prompts. The distribution of the survey responses to the five target prompts containing the scalar term *sometimes* is given in Table 4 below:²²

Table 4		Yes	No	I don't know
Category	Question #			
S	1	91.7%	8.3%	0.0%
TF	14	91.7%	4.2%	4.2%
U	4	45.8%	50.0%	4.2%
N	10	37.5%	58.3%	4.2%
	27	54.2%	45.8%	0.0%

Looking at this table, the first thing we notice is that the unstrengthened prompts containing *sometimes* also received a ~50/50 response. Thus, it turns out that the neutral prompts containing *sometimes* also acted in a similar fashion to the unstrengthened prompts containing the same scalar term (but not the same as the unstrengthened target prompts in general). In general the unstrengthened target prompts were only given a strengthened interpretation 33.7% of the time (see Table 3). In addition, we see that the strengthened target prompt containing the scalar term *sometimes* (Q1) received strengthened interpretations at a significantly higher rate than the strengthened prompt containing the scalar term *warm* ($p < .05$),²³ which was only strengthened 58.5% of the time.²⁴ In fact, Q1 was strengthened at the same frequency as the truth-functionally strengthened prompt containing *sometimes* (Q14), an unusually high rate of strengthening. Thus we see that across the board the target prompts containing *sometimes* generally received a higher frequency of strengthened responses as compared to the target prompts in the same category. The differences between *sometimes* and the other scalar terms (in particular, the differences between *some* and *sometimes*) will be discussed in more detail in Chapter IV.

Having discussed the neutral target prompts containing *sometimes* and *warm*, this leaves us with just the neutral target prompts containing *some* to analyze. Q16 (*If foxes kill some of the chickens, then we will have less eggs to sell*) fell into the ~50/50 category, while Q20 (*If some of the chickens are black, then there will be an announcement*) acted more like the unstrengthened target prompts. Looking at the different prompts, it makes some sense that Q16 would lean somewhat in the direction of a strengthened interpretation (Q16 was strengthened 57.7% of the time), whereas Q20 would lean more toward an unstrengthened interpretation (Q20 was given an unstrengthened interpretation

²² For details of this analysis see Analysis 2.11 in Appendix I.

²³ For details of this analysis see Analysis 2.12 in Appendix I.

²⁴ For details of this analysis see Analysis 2.13 in Appendix I.

69.2% of the time).²⁵ Q16 is unlike the rest of the neutral target prompts, in that there is a clear connection between the antecedent and the consequent. If there are less chickens because foxes kill some of the chickens, then it stands to reason that there would be less eggs to sell, because there will be less chickens laying eggs. In fact, it almost works like the truth-functionally strengthened target prompts, in that it only really makes sense given a strengthened interpretation. If foxes kill *all* the chickens then there won't be any eggs to sell. While this would technically be less eggs than before ($0 < n$ for $\forall n \in \mathbb{N}$), there is a reasonable resistance to interpreting this utterance that way, and a natural impulse to interpret the speaker as not intending or anticipating that all of the chickens will be killed by foxes, and thus the speaker means *if foxes kill some but not all of the chickens, then there will be less eggs to sell*. Obviously, there were a significant number of subjects who were willing to interpret "less eggs" as being possibly no eggs, and who thus assigned an unstrengthened interpretation to this prompt, which then explains why this prompt fell into the ~50/50 group. On the other hand, Q20 was more often assigned an unstrengthened interpretation, and again there is a fairly plausible explanation why this prompt would lean more in this direction. Q20 is a paradigmatic example of the neutral target prompts, in the sense that there is no clear connection between the antecedent and the consequent. Neither the given context or world knowledge tells us anything about why having black chickens is particularly announcement-worthy. However, given that the speaker (Jane) has made such a connection in her utterance, the natural assumption is that if having some black chickens is noteworthy, having more black chickens (all the chickens are black) is even more noteworthy, and so there is no reason to interpret Jane as meaning *if some but not all of the chickens are black, then there will be an announcement*. In conclusion, the neutral target prompts did turn out to be interesting, but not as uniform as the other categories of target prompts, and thus more difficult to interpret.

Thus far the reasoning processes in the interpretation of scalar implicatures in the antecedent on conditionals is basically a black box. We know that we generally get certain outputs (either strengthened or unstrengthened interpretations) for the different inputs (target and control prompts), but the survey data tell us nothing about how the subjects arrive at these interpretations. In order to look inside this black box we will need to examine the think-aloud data (Section II), and the connections between the think-aloud data and the survey responses (Section III).

Section II: Investigating the Data from the Think-Aloud Responses

The goal of the think-aloud portion of the experiment was to try and gain some insight into the answers to two additional questions:

1. What considerations do people appeal to when making decisions about scalar implicatures?
2. How do these considerations connect to the subject's decision to give a strengthened reading or an unstrengthened reading to a scalar term in the antecedent of a conditional?

²⁵ For details of this analysis see Analysis 2.9 in Appendix I.

Chapter III: Analysis of Experiment

In this section we will primarily be examining the answer to the first question, leaving the discussion of the second question for Section III. We begin by looking at the frequency of the occurrence of each of the think-aloud codes in response to the target and control prompts in order to see which codes occur most frequently in each type of prompt (target vs. control), and whether there are any interesting differences in the think-aloud responses of the 24 participants²⁶ to the different types of prompts. The codes under investigation are as follows:

- **H**: subject references a higher item on the scale (i.e., *all* or *most* rather than *some*, *hot* rather than *warm*, or *always* rather than *sometimes*).
- **L**: subject references a lower item on the scale (i.e., *few* or *none* rather than *some*, *cold* rather than *warm*, or *never* rather than *sometimes*).
- **A**: subject references alternate utterances the speaker might have made instead of the utterance given in the prompt
- **P**: subject considers possible alternate states of the world; i.e., subject references hypothetical alternate scenarios that might be (could have been) the case rather than the scenario given in the prompt
- **CS**: subject explicitly exhibits sensitivity to context, either by making reference to general world knowledge that he or she considers to be relevant in interpreting the speaker's utterance or by embellishing the context beyond what is given in the prompt. Alternatively, the subject explicitly asserts that he or she lacks information about the context or the speaker that is necessary for making an interpretation of the speaker's utterance.

As a first pass, we simply look to see how frequently each of these codes occurs in response to any prompt (without separating out the different types of prompts). The results of this analysis are given in Table 4 below:²⁷

Table 4	H	L	A	P	CS
	63.4%	15.7%	25.8%	24.7%	25.6%

As these results show, **H** was by far the most frequently used code with the other codes being used much less frequently. However, when we take a more nuanced look at the subjects' think-aloud responses by splitting the analysis into the target and control prompts and comparing the frequency of the occurrence of each code in response to the different types of prompts, a couple of interesting differences emerge. The results of this analysis are given in Table 5 below:²⁸

²⁶ Data from two subjects was removed from the analysis of the think-aloud responses because half of the think-aloud data for those subjects was lost due to equipment malfunction.

²⁷ For details of this analysis see Analysis 3.1 in Appendix I.

²⁸ For details of this analysis see Analysis 3.2 in Appendix I.

Table 5	Target	Control
H	61.7%	65.6%
L	15.6%	16.0%
A	12.5%	42.4%
P	40.3%	5.2%
CS	26.9%	24.0%

The frequency of response rates for the **H**, **L**, and **CS** codes in this analysis show little difference from the previous analysis (Table 4), since there is no statistically significant difference between the target and control prompts in terms of the frequency of these responses.²⁹ However, when we look at the response rates for the **A** and **P** responses, we *do* see a change. When the two types of prompts were analyzed together, **A** and **P** appeared with equal frequency (~25% of the time). However, when we analyze the target and control prompts separately, we see that there is a significant difference in the frequency of the occurrence of **A** and **P** responses in the target vs. the control items ($p < .001$),³⁰ with **A** responses occurring more frequently in the control prompts, and **P** responses occurring more frequently in the target prompts. Note the interesting flip-flop between the frequencies of these two codes in the target vs. the control prompts; **P** responses are given ~40% of the time in response to the target prompts, while **A** responses occur a corresponding amount of the time in the control prompts. Given this, and what we know about the frequency of **H** responses, two natural questions are as follows:

1. How often do **H** and **A** occur together in the control prompts?
2. How often do **H** and **P** occur together in the target prompts?

What we found was that within the control prompts that received an **A** response, 99.2% of those prompts also received an **H** response. Likewise, within the target prompts that received a **P** response, 93.1% of those prompts received an **H** response.³¹ Thus, it makes sense to include (**H&P**) and (**H&A**)³² as variables in further analysis where relevant. Thus, the previous analysis run on (**H&P**) and (**H&A**) gives the following result (see Table 6 below):³³

Table 6	Target	Control
H&P	36.9%	4.2%
H&A	10.8%	41.7%

As expected, the (**H&P**) results are virtually identical to the results of the same analysis run on **P**, as are the (**H&A**) results when compared with the **A** results (see Table 5).

²⁹ For details of this analysis see Analysis 3.3 in Appendix I.

³⁰ For details of this analysis see Analysis 3.3 in Appendix I.

³¹ For details of this analysis see Analysis 3.4 in Appendix I.

³² (**H&A**) = Boolean And of **H** and **A**; (**H&P**) = Boolean And of **H** and **P**

³³ For details of this analysis see Analysis 3.2 in Appendix I.

Thus far we have been looking at the survey responses (Section I) and the think-aloud responses (Section II) as if they were completely disconnected. However, recall that the goal with the think-aloud responses was to get a glimpse into the reasoning processes behind the derivation of scalar implicatures in the antecedent of conditionals, with the survey responses serving as a record of the output of this reasoning process. At this point, the reasoning process is still largely a black box, giving us an idea as to some of the considerations that are relevant to this process, but with no clear picture as to how these elements are connected to the output (the decision as to whether the utterance should be given a strengthened reading or not). We need to see what connections can be made between the think-aloud responses and the survey responses in order to open up this black box even further. We will explore these connections further in Section III.

Section III: Inside the Black Box: Connecting the Think-Aloud Responses and the Survey Responses

As a starting point, we begin by asking the following question: Separating the target prompts into those that the subjects assigned a strengthened interpretation ("Yes" response) vs. those that they assigned an unstrengthened reading ("No" response),³⁴ how often do subjects reference each of the various think-aloud responses (**H**, **L**, **H&A**, **H&P**, **CS**)? Secondly, how does this compare with the control prompts? Results of these analyses are given in Tables 7 and 8 below:³⁵

Table 7	Target	
	Yes (<i>n</i> = 191)	No (<i>n</i> = 151)
H	57.6%	68.2%
L	6.8%	25.2%
H&A	16.8%	4.0%
H&P	27.2%	49.7%
CS	28.8%	21.2%

Table 8	Control	
	Yes (<i>n</i> = 215)	No (<i>n</i> = 65)
H	78.1%	30.8%
L	3.3%	55.4%
H&A	55.8%	0.0%
H&P	4.2%	4.6%
CS	26.0%	13.8%

This level of analysis allows us to expose some differences between the target and control prompts that were not evident at the previous level of analysis (Tables 5 and 6). In the previous analysis, there was no significant difference between the target and the control prompts in terms of the frequency of **H**, **L**, or **CS** responses. However, looking at the results above, we see that the frequency of **H** or **L** responses is not in fact uniform across the different types of prompts. Recall that subjects gave an **H** response to 61.7% of the target and to 65.6% of the control prompts (see Table 5). Under the current analysis, the story for the target items remains much the same as before in terms of the frequency of **H** responses, since subjects gave **H** responses to both the items that they strengthened and to the items that they did not strengthen, although more frequently in the later (68.2% of the unstrengthened target items received an **H** response versus 57.6% in the strengthened target items). However, within the control prompts, we observe that

³⁴ Ignoring the "I don't know" responses as outliers

³⁵ For details of this analysis see Analysis 3.5 in Appendix I.

subjects gave **H** responses much more frequently in response to the items that they assigned a strengthened interpretation (78.1% of the control prompts that received a strengthened reading also received an **H** response), whereas they only give an **H** response to about a third (30.8%, $n = 20$) of the items that they assigned an unstrengthened interpretation.³⁶ The frequency of **L** responses above also shows some interesting variation between the items that received a "yes" response versus the items that received a "no" response. In the previous analysis (Table 5) we found that **L** responses occurred in response to 15.6% of the target prompts, and similarly, in response to 16.0% of the control prompts. In the current analysis we see that, in fact, **L** responses rarely occur in conjunction with a strengthened interpretation (in both the control and target prompts), but are relatively frequent in the items that were given an unstrengthened interpretation (25.2% of the unstrengthened target items received an **L** response, and 55.4% of the unstrengthened control items received an **L** response). However, it should also be noted that these **L** responses occurred most frequently in both target and control items that contained the scalar item *warm*.³⁷ We will discuss some of the interesting differences between *warm* and the other scalar terms further in Chapter IV.

Another interesting piece of the puzzle emerges when we look at the results for **H&A** and **H&P** responses. Recall that **H&A** is the conjunction of the **H** responses and the **A** responses, and **H&P** is the conjunction of the **H** responses and the **P** responses. Basically, **H&A** is the response pattern that corresponds to the "textbook" story about scalar inferences; that is, the speaker could have said the higher item on the scale (e.g., they could have said *all* rather than *some*), but they didn't, so they must have meant *some but not all*. Falling nicely in line with this story, we see that in the control prompts, 100% of the items that received **H&A** responses also correspond to items that received a strengthened reading.³⁸ Although only a small number of the target prompts received **H&A** responses ($n = 39$),³⁹ the majority of those responses also corresponded to target prompts that received a strengthened reading (82.1%),⁴⁰ so the subjects' responses to the target prompts also appears to follow the standard scalar story, at least to the extent that when subjects do consider alternate utterances containing higher items on the scale that the speaker might have said but didn't, they also generally assign a strengthened reading to the speaker's utterance. However, as we have observed, **H&A** responses are relatively infrequent in reference to the target prompts, since they occurred in only 10.8% of the

³⁶ Recall also that most of these items contain the scalar term *warm* (see Table 2), which we have already observed makes them behave differently than the other two scalars.

³⁷ Of the target items that received **L** responses, 68.6% of them contained the scalar term *warm*, and 90.7% of the control items that received an **L** response were items that contained *warm*. For details of this analysis see Analysis 3.6 in Appendix I.

³⁸ Although it is worth noting that **H&A** does not seem to have been the only relevant consideration, since **H&A** responses appeared in only 55% of the control items that received a strengthened interpretation. **CS** was also a relatively frequent response, occurring in 26.0% of the control items that received a strengthened interpretation, overlapping to some degree ($n = 17$) with the **H&A** responses (see Analysis 3.7 in Appendix I for details).

³⁹ See Analysis 3.2 in Appendix I.

⁴⁰ For details of this analysis see Analysis 3.8 in Appendix I.

targets,⁴¹ as compared to 41.7% of the controls (see Table 6). Instead, the target prompts appear to receive **H&P** responses with a frequency that corresponds closely to the frequency of **H&A** responses in the control prompts. This suggests **H&P** as a possible reasoning pattern for the target items (conditionals), analogous to the **H&A** pattern of reasoning that we see so strongly in the control items. This has some plausibility, since in many cases it does not make sense to use the **H&A** reasoning in conditionals. Consider the following example:

- (14) Two TAs are discussing their students' upcoming assignments. One TA says to the other:
If some of my students turn their papers in late then it throws my whole schedule off.

This example seems to lean toward an unstrengthened reading. If all of the students turn their papers in late, then it will definitely throw the TA's schedule off, so it seems clear that what the TA means in this case is *If some and possibly all of my students turn their papers in late then it throws my whole schedule off*. If we try to use the (**H&A**) reasoning process on this utterance (e.g., the TA could have said *If all of my students turn their papers in late then it throws my whole schedule off*, but she didn't. Therefore, she must have meant *If some but not all of my students turn their papers in late then it throws my whole schedule off*), we get results that seem intuitively incorrect. Moreover, the reasoning process seems off. The higher item on the scale seems like too high a threshold for most hypothetical situations, and therefore it is perhaps less likely to be uttered in the antecedent of a conditional. Since antecedents are sufficient conditions for their consequents, the lower item on the scale (e.g., *some*) seems very natural as setting a lower bound on the conditions that will serve to activate the consequent, even in cases where the meaning is *some (and possibly all)*, reserving the use of the higher items on the scale for more rare cases where you require that the higher item holds (e.g., *all of the student fail the exam*) in order for some consequence to follow (e.g., the professor will give a new exam). The **H&P** response pattern, on the other hand, seems like a much more natural way to reason through the interpretation of conditionals.

To illustrate how this pattern of reasoning goes, consider the following prompt from the experiment, along with a sample of some of the **H&P** responses that subjects gave in response to this item during the think-aloud portion of the experiment:

30. Laura says to her mother:
"If it is warm outside, then I don't need a sweater."
Does Laura mean:
If it is warm but not hot outside, then I don't need a sweater.

Yes
 No
 I don't know

⁴¹ Interestingly, this response occurred most frequently ($n = 18$) in response to the "neutral" target items. For details of this analysis see Analysis 3.9 in Appendix I.

Chapter III: Analysis of Experiment

- Subject 101: I'm going to say "no," because if it was hot outside she certainly wouldn't need a sweater. (subject's survey response = no)
- Subject 108: Again, *warm* and *hot* are - in this sense more or less refer to the same thing – if it's a certain temperature outside, then she doesn't need a sweater, so warm or hot, wouldn't really make a difference. If anything, if it was hot it would be more incentive to not wear a sweater, so I would say "no." (subject's survey response = no)
- Subject 114: Uh, no, I don't think Laura means that. If it was hot, she wouldn't need a sweater either, um, so, I'm going to say "no." (subject's survey response = no)
- Subject 127: So, uh, no, because what Laura is trying to do now is ... stay cool, um, if it's warm outside then it's probably like a comfortable temperature to wearing like a t-shirt or something, so she's saying she doesn't need anything warmer than that, but if it's hot outside she also doesn't need anything to make her warmer ... um, so whether or not it's warm or hot – either way she doesn't need to be warmed up by a sweater. (subject's survey response = no)

As can be seen from the above think-aloud responses, the **H&P** response pattern involves the subjects checking their intuitions about whether or not the consequent would still follow if the higher item on the scale was the case (e.g., if it was hot outside rather than just warm). Unsurprisingly, we found that in case of the question above, subjects were fairly unanimous in deciding that no, Laura still did not need a sweater, even when it was hot outside. Basically, the idea seemed to be that if it made no difference to the consequent whether the lower item or the higher item on the scale was the case, then the speaker did not intend the scalar item to have a strengthened reading. In cases where it did make a difference to the consequent, subjects interpreted the scalar item as having a strengthened reading. We illustrate this behavior with some of the responses to another one of the target items which had a very high rate of **H&P** responses (83.3%). This prompt was very interesting because it elicited both strengthened and unstrengthened interpretations, depending on how subjects interpreted the context, thus serving as a good illustration of how subjects use decisions about whether the consequent will still follow under alternate scenarios where the higher item on the scale is the case (all of the students fail the exam, rather than just some of them) in order to decide whether the speaker intended the scalar item to have a strengthened interpretation or not.

18. A professor and his TA are having a meeting. The professor says:

"If some of the students fail the exam, then I will give an extra credit assignment."

Does the professor mean:

If some but not all of his students fail the exam, then I will give an extra credit assignment.

- Yes
- No
- I don't know

- Subject 102: Yes, 'cause ... maybe not. Okay, so, if some of them [fail the exam], you'll give extra credit, but if all of them ... he might do something else. Okay, so yeah, it is saying the same thing. (subject's survey response = yes)
- Subject 105: Well, I think that if all of the students failed then he would also give an extra credit assignment, so I'll disagree. If *some*, or *all* – it's the same idea here. (subject's survey response = no)
- Subject 106: No, he means that ... no. Because if some, or more fail, then he will give a credit assignment. So it's not only if some fail, because if all fail then he'll have to give a credit assignment. (subject's survey response = no)
- Subject 113: Um ... yes, because, he would be concerned ... yes. He's saying – he means that *some but not all*, because if all the students failed, he probably would just have curved the exam, and done something completely different, because obviously that would mean a whole fail instead of just a few kids who would need the extra credit – and then a bunch of other kids, like the ones who did well, who would not need it. So, yes, he means *some but not all*, because if all of them he would have done something completely different and said something different. (subject's survey response = yes)⁴²
- Subject 118: Um ... hmmm, interesting, because I would read that to mean *I will give an extra credit assignment to those of the students who fail the exam, if there are any of them*. So he would still give the assignment to the students who failed the exam if all of them did – he'd give it to all of them. Um, and I guess if none of the students failed, he'd have no need to devise the assignment in the first place. Um, so, I don't think there'd be any real difference between whether some of them or all of them failed, so I'll click "no." If all of the students fail the exam, he'll still give an extra credit assignment. (subject's survey response = no)

As you can see from these sample responses, whether the subjects gave a strengthened interpretation (survey response = yes) or an unstrengthened interpretation (survey response = no) pretty much depended on what they thought would happen if all the students failed the exam. Those that thought that the professor would give the extra credit assignment if either just some, or all, the students failed the exam assigned an unstrengthened interpretation, while those that thought that the professor would do something else if all of the students failed assigned a strengthened reading. In this case both interpretations seem plausible, given the right set of background assumptions, unlike the previous example, where world knowledge practically guarantees an unstrengthened interpretation. The overall point here is that the **H&P** pattern of reasoning can plausibly have as an output either a strengthened or unstrengthened interpretation, depending on the context (including the content of the conditional as a crucial piece of the context!). This differs from the **H&A** pattern of reasoning, which generally seems to output only strengthened interpretations (in both the target and control items). Referring back to Table 7, we see that the **H&P** response was used by subjects in 27.2% of the target prompts that received a strengthened interpretation, and in 49.7% of those that received an unstrengthened interpretation. There are several possible explanations for the higher rate of **H&P** responses in the target items that received unstrengthened interpretations:

⁴² Note: this subject's think-aloud response was coded as **H&A** as well as **H&P**.

Chapter III: Analysis of Experiment

1. Using the **H&P** pattern of reasoning causes subjects to give an unstrengthened interpretation to the scalar item in the antecedent of a conditional.
2. Conditionals that lean toward an unstrengthened interpretation lend themselves to the **H&P** reading more so than conditionals that lean toward a strengthened interpretation.
3. Given the way the prompts in this experiment were set up (the strengthened interpretation was given as part of the prompt), subjects needed to "reach" more to arrive at an unstrengthened reading and thus were more likely feel the need to "walk through" the reasoning process explicitly, thus increasing the rate of **H&P** responses for those items.

Explanation (1) seems clearly false, since **H&P** responses occurred in target prompts that received either a strengthened or unstrengthened interpretation, and, as we showed in the examples above, can seem very natural and plausible in both cases. Explanation (2) is possible; however, given the fact that **H&P** responses *do* occur in a reasonable number of the target items that received a strengthened interpretation (27.2%), explanation (3) seems like the more reasonable possible explanation. Teasing apart these two possibilities (explanations 2 and 3) is an avenue that should be explored in further experimental work by presenting subjects with some prompts that gave a strengthened reading as a potential interpretation, and other prompts that gave an unstrengthened reading as the potential interpretation. Crucially, some of the prompts that supply the unstrengthened reading as a potential interpretation should be ones that seem to lean in the other direction (i.e., lean toward a strengthened interpretation), causing subjects to have to "reach" for the strengthened as well as unstrengthened interpretations, thus encouraging more explicit reasoning processes in the both categories of target prompts (if explanation 3 is correct).

Another possible explanation is that the "Yes/No" answers are not really about the subjects' interpretation of the conditional itself, but rather reflect some further conclusions that they are inclined to draw. This is a general difficulty with the experimental paradigms, in that it is difficult to design prompts such that you can effectively examine the phenomenon in question, without biasing the results. Geurts & Pouscoulous (2009) discussed this issue in relation to their experiments. They have argued elsewhere (Pouscoulous 2006, Geurts 2009) that the introspective method is biased when it comes to judgments about scalar implicatures. Geurts (2009) illustrates the introspective method with the following example:

- (15) a. Some of the goats have a cough.
b. Not all of the goats have a cough.

The introspective method is biased according to Geurts (2009) for the following reason:

Obviously, to ask oneself whether or not [(15a)] implies [(15b)] is to suggest already that it might be implied, but more importantly, this question raises the issue whether or not all of the goats have a cough, or in other words, it makes it relevant to establish whether or not this is the case (p. 61).

The problem is that the experimental paradigm that they (and we) used basically comes down to collecting introspective judgments on scalar implicatures from a population of naïve native speakers. Thus, the experimental method inherits the problems of the introspective method, in that the experimental prompts make the scalar implicature relevant in a way it might not otherwise have been, potentially causing an inflated rate of endorsed scalar implicatures. This was a worry for Geurts & Pouscoulous' in their experiment, and should not be lightly dismissed in terms of our experiment, which is modeled on theirs. We tried to correct for this as much as possible in designing our experimental prompts, through allowing subjects the option of answering "I don't know" (as opposed to just "Yes" or "No") in response to the proposed scalar implicature, and also in asking the subjects to freely generate their own explanations of their reasoning process in arriving at their interpretations of the experimental prompts. Obviously, this does not do away with any bias introduced by making the potential scalar implicature salient in a way it might otherwise have been if the prompt had been given without the potential scalar implicature. However, if we had presented the prompt in a more open-ended fashion as in (3) below, allowing the subjects to freely generate the scalar implicatures (or not), this would also have created problems for the experiment.

(6) Bob is telling his wife about his performance at the Harvest Festival. He says:

I danced some of the dances.

What does Bob mean by his statement?

Leaving the question of what Bob means by his utterance of *I danced some of the dances* open-ended in this fashion would invariably have introduced a lot of noise into the data, not to mention making the data harder to analyze. As we saw in analyzing the think-aloud portion of the experiment, standardizing subjects' free responses in such a way that they can be easily compared is both difficult and time-consuming. Thus, while asking the subjects if Bob meant *I danced some but not all of the dances* introduces some bias into the experiment, doing so was a practical necessity in order to get standardized and easily analyzable results.

Additionally, it is not clear that leaving the question as a free response on the part of the subjects would have eliminated the bias, since there is an inherent bias possibility in subjects' free responses, in that they want to please the experimenter and thus try to figure out what the "right" answer is. This is an issue that is relevant for the think-aloud portion of the experiment. We tried to correct for this as much as possible by giving the instructions for the think-aloud portion of the experiment in as open-ended a way as possible, and by not giving them any feedback on their responses, as much as was humanly possible. However, it is clear from subjects' think-aloud responses that there was learning going on in terms of subjects' responses to the experimental prompts, and that they did tend to settle into something of a groove in their responses, although they could be jolted out of that groove by different prompts. Thus, we need to be careful about drawing too strong a conclusion about subjects' reaction to the experimental prompts, either in terms of their interpretations of the scalar terms or in terms of the reasoning patterns that we see them

employing to arrive at these interpretations. However, as Geurts & Pouscoulous (2009) conclude, these potential biases do not mean that the experimental paradigm is irredeemably flawed. They merely indicate a need for caution in interpreting the results.

Section IV: Conclusions

Although more work remains to be done in order to fully understand the phenomenon of scalar implicature in conditionals, the data collected in this experiment has enabled us to answer a number of questions about the phenomenon, and given us an intriguing glimpse into the pragmatic reasoning processes that people use to arrive at interpretations of scalar implicatures in the antecedent of conditionals, as well as giving us more insight into the behavior of unembedded scalar implicatures. To sum up, in Section I we answered two basic questions:

1. Do scalar implicatures ever show up in the antecedent of conditionals?
2. Do they always show up in the antecedent of conditionals?

The answer to the first question was, clearly, yes, scalar terms in the antecedent of conditionals *do* receive strengthened interpretations with reasonable frequency. The answer to the second question was equally clear: no, scalar terms in the antecedent of conditionals do not *always* receive a strengthened interpretation. This clearly differed from the unembedded control items, which robustly received a strengthened interpretation (especially if you set aside the control prompts with *warm*, which, as we have observed, behaves differently from the other scalar terms included in this experiment). This result serves to confirm the intuitions discussed in Chapter I, that scalar terms are sometimes strengthened in the antecedent of a conditional, and sometimes not. In addition, we predicted that certain of the target prompts would lean toward a strengthened interpretation, while others would lean toward an unstrengthened interpretation, and these predictions were also confirmed by the survey response data.

Using the data from the think-aloud responses, we were able to investigate the considerations that subjects expressed as relevant to their decision as to whether scalar terms (both in unembedded contexts, and in the antecedent of conditionals) should receive a strengthened or unstrengthened interpretation. We found some similarities and some differences between the subjects' responses to the embedded and unembedded scalar implicatures. Reference to a higher item on the scale (**H**) was common in both the target prompts (conditionals) and the control prompts (unembedded); however, there was a difference in the way the higher item on the scale was referenced in the different types of prompts. In the control prompts the higher item on the scale was referenced as an alternate utterance that the speaker could have made but didn't (**H&A**). In the target prompts the higher item on the scale was referenced not as an alternate *utterance* that the speaker might have produced, but rather as an alternate (hypothetical) way the world might be in terms of the antecedent of the conditional, in order to check intuitions about whether or not the speaker would intend the consequent to still follow under this hypothetical antecedent (**H&P**). Based on these intuitions, subjects then either chose a strengthened interpretation (if they determined that the consequent would not follow

under the higher item on the scale) or an unstrengthened interpretation (if they decided that the consequent would still happen even when the higher item on the scale was substituted for the lower item).

This difference in the reasoning patterns in the control and target prompts (**H&A** in the control prompts, **H&P** in the target prompts) was the primary insight that was gleaned from the think-aloud responses, and has been the primary focus of the discussion in this section thus far; however, some attention should be paid to other insights that can be gleaned from the subjects' think-aloud responses. In addition to references to higher items in the scale, whether in reference to alternate utterance (**H&A**) or to alternate possibilities (**H&P**), subjects also referenced "lower" items on the scale; that is, subjects made reference to *none* rather than *some*, *never* rather than *sometimes*, or *cold* rather than *warm*. In particular, subjects primarily referenced *cold* in reference to *warm* (~80% of the **L** responses were in reference to prompts that contained the scalar term *warm*).⁴³ This is not the only difference between *warm* and the other scalar terms under investigation. Throughout this analysis, we have observed that the different scalar terms do not behave in a uniform fashion, with *warm* being the most obviously anomalous. This is not an issue that we originally set out to investigate, but it is obviously a finding worth pursuing in as much depth as our data will allow. In Chapter IV we will explore the differences between the different scalar terms (*some*, *sometimes*, and *warm*) that emerged during the course of our analysis.

⁴³ For details of this analysis see Analysis 3.10 in Appendix I.

Section I: *Warm* is Less Robust Than *Some* or *Sometimes*

As we noted in the previous sections, the prompts containing the scalar term *warm* did not elicit the same reactions as the prompts with the other two scalars. We first noted this in Section I when we examined the distribution of the survey responses to the target and control prompts broken down by scalar term, and observed that the behavior of the control prompts containing the scalar term *warm* differed markedly from the control prompts with *some* or *sometimes* (see Table 2). In the control prompts containing *some* and *sometimes* we found the results that we were expecting in the control prompts: over 90% of these prompts received a strengthened reading. We found a different story in the control prompts containing the scalar term *warm*: these prompts were only given a strengthened interpretation ~35% of the time – in fact, they were given an unstrengthened interpretation more often than not (58.7% of the control prompts containing the scalar term *warm* were given an unstrengthened interpretation). Obviously, this goes against the standard scalar story. According to Horn, *hot* and *warm* are scalar alternatives in the same way that *all* and *some* are. If this is correct, then *warm* should behave the same as *some*; however, this is clearly not the case, as we can see from the discussion above.

We gained further insight into the differences between the different scalar terms by looking at the data from the think-aloud portion of the experiment. Recall the results that we found when we compared the frequency of the **H** and **L** responses to the target and control prompts (see Table 5, reproduced below in part as Table 9):

Table 9	Target	Control
H	61.7%	65.6%
L	15.6%	16.0%

If you break this analysis down by scalar, some interesting differences emerge in the frequency of the **H** and **L** responses for the control and target prompts across the different scalar terms (see Table 10 below).⁴⁴

Table 10		Target	Control
H	<i>some</i>	75.0%	84.4%
	<i>sometimes</i>	41.7%	61.5%
	<i>warm</i>	68.3%	51.0%
L	<i>some</i>	8.3%	4.2%
	<i>sometimes</i>	5.0%	0.0%
	<i>warm</i>	33.3%	43.8%

Looking at these results, we see that there are some differences⁴⁵ in the frequency of **H** responses to the different types of prompts. However, in all cases **H** responses occurred

⁴⁴ For details of this analysis see Analysis 3.11 in Appendix I.

⁴⁵ Specifically, there is a significant difference between *some* and *sometimes* and *some* and *warm* in the control prompts ($p < .001$), and within the target prompts there is a significant difference between *some* and *sometimes* and *warm* and *sometimes* ($p < .001$). For details of this analysis see Analysis 3.12 in Appendix II.

with reasonable frequency, and tell much the same story as we saw at the previous level of analysis (Table 9). In fact, the most significant thing that jumps out is the difference in the frequency of **L** responses in the prompts containing the scalar term *warm* versus the prompts containing the scalar terms *some* or *sometimes* (both in the target and control prompts). Subjects rarely gave an **L** response to the prompts containing *some* or *sometimes*, but did so significantly more often in response to the prompts containing the scalar term *warm* ($p < .001$).⁴⁶ This goes along with the result we discussed at the beginning of Section III (see Tables 7 and 8), where we observed that **L** responses occurred much more frequently in the prompts (both target and control) that received an unstrengthened interpretation. Non-coincidentally, more of the prompts that received an unstrengthened interpretation contained the scalar term *warm*. This was especially true in the control prompts, where the difference between the scalars was marked (58.7% of the control prompts containing *warm* were given an unstrengthened interpretation, as compared to only 1.9% of the prompts containing *some* or 6.7% of the prompts containing *sometimes* – see Table 2). This makes a certain amount of sense, because you would expect that reference to a "lower" item on the scale (**L**) would incline the subject toward an unstrengthened reading,⁴⁷ since an **L** response indicates that it is the comparison between *warm* and *cold* that is salient for the subject in this situation. Thus whether it is warm or hot is not taken to be important to the speaker and thus the speaker did not intend *warm* to be interpreted as *warm but not hot*. Additionally, since **L** responses were much more common in reference to prompts containing the scalar term *warm*, and these items were often given an unstrengthened interpretation, it makes sense that all of these things would be clustered together. To see how all of these elements come into play in the subjects' think-aloud responses, we look at some samples of **L** responses from subjects to prompts containing the scalar term *warm*. We start with a few unembedded (control) examples:

19. The web page on how to stay warm in the winter gives the following tip:

"Wearing fleece slippers around the house can help keep your feet warm."

Does the web page mean:

Wearing fleece slippers around the house can help keep your feet warm but not hot.

Yes

No

I don't know

- Subject 101: I'm saying "no," for the same reason as the pie one [Q7]. (Investigator: Which is? Just could you say it again?) Which is that it's warm as opposed to cold, not which degree of warm. (survey response = no)

⁴⁶ For details of this analysis see Analysis 3.12 in Appendix I.

⁴⁷ Specifically, 72.5% of the prompts that received an **L** response were given an unstrengthened interpretation. For details of this analysis see Analysis 3.13 in Appendix I.

Chapter IV: Further Results: Behavior of Scalar Terms is not Uniform

- Subject 102: Um ... I think .. for this one it's saying ... it'll keep your feet warm but not cold, because it's a web page on how to stay warm in the winter, and not ... how to stay away from being hot – even though they're the same thing it's like... the point of view is different, so it could be like, conveying a different message, so ... (survey response = no)
- Subject 113: No, they're not concerned about specific temperatures – they're just saying if you wear slippers you're not going to be cold. (survey response = no)
- Subject 118: Um ... I – I guess so? Um, I think really what they want to say is they'll keep your feet warm, rather than cold... uh, not too sure people wondering how to stay warm in winter are afraid of being hot in winter ... um, so I'm going to go with "no," that's not really the implication they're after. (survey response = no)
- Subject 124: No. It's like the ... afghan sentence from earlier [Q9] .. the default condition is your feet are cold, so wearing these fleece slippers can help them to be warm, as opposed to cold, not warm as opposed to hot. (survey response = no)

As these examples show, the contrast that the subjects were focusing on was a *warm/cold* contrast, rather than a *warm/hot* contrast. In this example the subjects mainly focused on the desire to have your feet be warm rather than cold, and thus the possibility of your feet being hot was taken as either not being a huge concern, or not likely to happen. Thus, based on this assessment of the unlikelihood or unconcern about the feet being hot, the subjects concluded that the speaker did not mean *warm but not hot*. This motivation was a common theme in almost all of the control prompts containing the scalar term *warm*, possibly because of commonalities between the background contexts of these particular prompts (a large number of the prompts unintentionally involved something about winter days, or night time, when people are presumably more worried about being too cold). Thus, while we note with interest that the salient alternative to *warm* seems to be *cold* rather than *hot*, we also want to be cautious about drawing too strong a conclusion on the basis of these incidental results from our experiment. More focused experiments would need to be run in order to come to more definitive answers about the differences between the behaviors of the different scalar items. For now, all we can do is point out some of the interesting and possibly significant features of the various responses to the prompts containing *warm*. In line with that goal, we next highlight another possible reason to reject the strengthened reading of *warm*, which came up in a few of the subjects' responses to this prompt:

9. Kathy has crocheted an afghan for her new niece Alice. Kathy says:

"This blanket will keep Alice warm at night."

Does Kathy mean:

This blanket will keep Alice warm but not hot at night.

Yes

No

I don't know

Chapter IV: Further Results: Behavior of Scalar Terms is not Uniform

- Subject 102: Noooo, because she doesn't know .. like – she might not know the warmth of the blanket, and, um I think by saying that it's just like – saying *the blanket will keep Alice warm at night* ... that ... she [Alice] won't be cold, instead of not hot, so no. (survey response = no)
- Subject 111: Nooo. I don't think she can guarantee that [Alice will be warm but not hot]. I think she means ... when Alice needs to not be cold, um this blanket will keep her warm. Hmm she could be saying that it's just enough heat – it's not too heavy, but, um ... I don't think she's saying that. I don't think that you can make that conclusion, based on what she says. (survey response = no)

Here, the subjects are invoking Kathy's lack of knowledge or control over whether her niece will get hot or not. In other words, since Kathy cannot guarantee that the blanket will keep Alice warm but not hot, she cannot mean that, either. Likewise, we can see subjects responding in a similar fashion to the prompt below:

28. Cindy is putting on her favorite blue sweater. She says to her mother:
"This sweater keeps me warm on winter days."
Does Cindy mean:
This sweater keeps me warm but not hot on winter days.
 Yes
 No
 I don't know

- Subject 102: Um not necessarily, because a sweater could make you hot during the day if ... like, the winter day is a warmer one, so um, I would say "no" to this. Again, it's like not saying *not hot*, it's saying *not cold*. (survey response = no)
- Subject 111: I don't not necessarily. Um She could become hot, but she won't be cold. Um ... so, I don't know. I mean, I think no, but *This sweater keeps me warm but not hot on winter days* ... that's not what she means. She just means that it keeps her not cold, but it might keep her hot.
- Subject 127: Okay, again ... um, because it's the winter she's not uh ... I mean, because it's the winter her main goal is to not be cold. Uh, and because she'll probably still have on a coat or something – depending on the climate and stuff – but um, assuming it's like, super-cold she'll probably have more layers on over her sweater, so she's not concerned about being too hot – she's concerned about being too cold ... um... so, again, I think it's just getting away from the negative state that is coldness, so it could be warm or hot – I mean, hot wouldn't be ideal again, but I don't think she's really discriminating between warm and hot. (survey response = no)

Again, we see that the possibility that the sweater might make Cindy hot in the right circumstances leads these subjects to conclude that she does not mean *warm but not hot*.

Combined with this we also see the first sort of reasons that were discussed, where the possibility of Cindy being cold is more salient for the subject than the possibility of her being hot, and thus the subject concludes that she does not mean *warm* in the strengthened sense.

As we have seen just by looking at the *warm* control prompts, subjects' analysis of the meaning of these prompts involves a variety of contextual factors. The above prompt (Q28) was the control prompt that worked the best, in terms of eliciting the strengthened response predicted for the control prompts (Q28 was strengthened 42.3% of the time, as opposed to the other control prompts containing *warm*, which were only strengthened ~30-35% of the time).⁴⁸ However, the factor that seemed to tip the balance and make this prompt more likely to be strengthened than the other control prompts involving *warm* was the information that the sweater was Cindy's favorite sweater.⁴⁹ This was an unintended result (the part about the sweater being Cindy's favorite was not inserted intentionally to elicit this response), and shows further how *warm* is affected by a wide variety of contextual factors, in a way that *some* and *sometimes*⁵⁰ seem not to be.

Turning to the target prompts, we see that they exhibit a similar pattern of **L** responses as the control prompts; **L** responses rarely occur in response to the target prompts containing *some* or *sometimes*, and frequently occur in response to the target prompts containing *warm*. For an example of how these **L** responses come into play in the target prompts containing *warm*, we look at some sample responses to the prompt below, which we had looked at previously in the context of the **H&P** responses (Chapter III, Section III):

30. Laura says to her mother:

"If it is warm outside, then I don't need a sweater."

Does Laura mean:

If it is warm but not hot outside, then I don't need a sweater.

Yes

No

I don't know

- Subject 116: Um, I think ... no. she meant if it is *warm, but not cold outside, then I don't need a sweater* ... which makes sense, yeah, so I'm going to go with "no," she didn't mean that. (survey response = no)
- Subject 118: No, because if it was hot she also wouldn't need a sweater. So, she means if it is *warm* rather than *cold*. (survey response = no)

⁴⁸ For details of this analysis see Analysis 2.14 in Appendix I.

⁴⁹ Specifically, 50% of the people who responded "Yes" to this prompt mentioned something about the sweater being Cathy's favorite. For details of this analysis see Analysis 3.14 in Appendix I.

⁵⁰ *Sometimes* is also less robust than *some*; however, both *some* and *sometimes* were relatively immune to effects of context, at least in the control cases. Other issues pertaining to *sometimes* will be discussed later in Section II.

Chapter IV: Further Results: Behavior of Scalar Terms is not Uniform

- Subject 123: No, she just means if it's not cold – like, she would – if it's not cold, then she doesn't need a sweater, so, um, I mean, *hot* would be under *not cold*, so she wouldn't need a sweater then either. (survey response = no)
- Subject 124: Well, she doesn't need a sweater in either occasion [either warm or hot]. She's saying *if it's warm, not cold, then I don't need a sweater*. She would need a sweater if it was cold. (survey response = no)

As in the control prompts containing *warm*, we see *cold* being straightforwardly taken as the salient alternative to *warm*, thus prompting an unstrengthened interpretation (for example, consider Subject 116's response). However, we also see subjects giving an **L** response along with the corresponding unstrengthened interpretation, based on the **H** response not being a good "fit" for the situation. For example, Subject 118 first considers what would happen in a situation where the higher alternative on the scale was the case (if it was hot outside rather than just warm) and determines that the consequent would remain the same in that situation (Laura still wouldn't need a sweater). This is an example of the **H&P** reasoning pattern that was discussed in Chapter III, Section III; however, in the examples of the think-aloud responses discussed there, the speaker's meaning was either explicitly or implicitly taken by the subject to be *warm and possibly hot* or, more commonly, *some and possibly all* (given that most of the examples discussed in Chapter III, Section III involved the scalar term *some* rather than *sometimes* or *warm*). In contrast, we see here that subjects either implicitly or explicitly understand the speaker to mean *warm but not cold*, rather than (or perhaps in addition to) *warm and possibly hot*. Subjects take the speaker's utterance of *warm* in this context to be signifying a lower bound on the temperature conditions in the utterance (*warm but not cold*) rather than an upper bound (*warm but not hot*), and thus they respond "no" to the *warm but not hot* interpretation given in the prompt.

In other cases the speaker is taken as using *warm* in order to in some sense signify both a lower and an upper bound on the desired temperature. Consider some of the responses to the prompt given below:

36. Jim lives in England with his wife and two daughters. He and his wife are planning a trip to visit his sister in Pennsylvania. Jim says:

"If we want it to be warm when we visit, we should go in April."

Does Jim mean:

If we want it to be warm but not hot when we visit, we should go in April.

Yes

No

I don't know

- Subject 101: I'm actually going to say "yes" to this, um, because if they went in, say, June it would probably be hot, so he may actually mean *if we want it to be warm but not hot nor cold*. (survey response = yes)

Chapter IV: Further Results: Behavior of Scalar Terms is not Uniform

- Subject 115: Um, yes, because he's specifying April, he's not saying um, *if we want it be not cold when we visit, then we should visit in April*, because if he wants – if he just cares – if he only cares about the cold weather then he would say *we should go in between April and October*, but he's specifying April, and April is warm – it's neither cold nor hot, so he's caring about – he only wants warm, he doesn't want hot or cold. (survey response = yes)
- Subject 118: Um ... well, because he's talking about April ... uh, probably yes. If they wanted it to be hot they could go at the height of summer. Um,...yeah. I'm not sure how warm April is, um, but yeah, they could definitely go for a hotter month, so *warm but not hot* is a possibility, *warm rather than cold* is also a possibility – they could go in a much colder month than April ... um, so there's no reason for him to not mean that. He might mean something else as well – I don't know – but yeah, *warm but not hot* would make sense. (survey response = yes)

From these responses we can see that the choice of month was seen as significant in deciding what the speaker meant by *warm* in this utterance. By choosing a month that was at least likely to be warm rather than cold but also not likely to be too hot Jim was interpreted as being concerned with avoiding both extremes. However, a good deal of the strength of this effect depended whether the subject interpreted Jim's desire to go in April as based on an avoidance of cold, as opposed to wanting the perfect temperature. Consider a few more of the sample responses to this prompt, where subjects also seem to be considering the question of whether *warm* is meant as placing both an upper bound and a lower bound on the desired temperature; however, unlike the previous cases, the subject ultimately interprets *warm* as simply *warm but not cold* and thus responds in the negative to the proposed *warm but not hot* interpretation.

- Subject 102: Um...I think that ... he's saying that *if we want it to be warm, as opposed to cold, we should go in April*, because um well, it could go either way, because it gets really hot in Pittsburgh, and it gets really cold in Pittsburgh. And...I don't really know what it's like in England. So ... hmmm. Because he could also, like, have suggested like, uh, July, and it would be really hot, so it could be either way. But, automatically, I would assume Pittsburgh and cold, and they want to be in the warm, and um, I would say "no." (survey response = no)
- Subject 106: No ... no, because uh ... he doesn't want to be cold. Right. So he's saying that *oh, we should go when it's warm*. So he – he's not saying that *oh, we should not go because it's going to be hot otherwise, so we go in April*. Instead he's saying *we're going to be cold otherwise, so let's go when it's warm*. (survey response = no)
- Subject 113: No. He's not necessarily excluding anything. It could be hot – that'd probably even better thing, because it's always cold in Pennsylvania, and terrible. So *if we want it to be warm and not hot, we should go in April* – no, he's not excluding any particular temperature by saying he wants to go in April. (survey response = no)

- Subject 114: Um ... I think.. again, they – if they want to be warm, they're not thinking *all right, we want to be warm, but not too hot!* – they're just saying that – from what he says I feel like they're just worrying about being cold. They're not really even considering being too hot. Uh, so I'm going to say I think that Jim *if we want to be warm but not hot ...* I don't think that's what he's saying. I don't think – again, *hot* has really much to do with it, because he's just saying *warm ...* I mean, that's just what I'm gathering, and I'm hearing Jim say this, and I don't think he's saying *we're going to be warm, but not too hot if we go in April.* I think he just means that they're going to be warmer than if we were to go in February, or whatever. So, I'm going to say "no." (survey response = no)

In these responses we can also see how other contextual factors come into play in response to this prompt, such as the subject's perception of how cold⁵¹ it is in Pennsylvania during various months of the year.

In conclusion, we have seen how context or domain specific information has a very strong effect on the prompts containing the scalar term *warm*, both in the target and control prompts. In contrast, prompts containing the scalar terms *some* or *sometimes* were affected by context or domain specific information in the target prompts, but the control prompts showed little or no effect of context or domain specific information. Recall the results given in Table 2 (Chapter III, Section I) for the control prompts that first drew our attention to the differences between the different scalar items. Control prompts with *some* and *sometimes* nearly always received a strengthened interpretation (96.2% and 91.3%, respectively), whereas control prompts with *warm* were strengthened significantly less often (only 35.6%). In addition, a closer examination of the reasoning patterns that were identified in conjunction with the different types of prompts provide us with further evidence of the non-uniform behavior of the different scalar items. Recall the results given in Table 6, reproduced below as Table 11:

Table 11	Target	Control
H&P	36.9%	4.2%
H&A	10.8%	41.7%

As we can see, **H&A** was used most frequently with the control items, and that **H&P** was used correspondingly often with the target items. However, the differences within the

⁵¹ The focus on how cold it is in Pennsylvania in the responses to this prompt may have been somewhat affected by real-time contextual factors, since the data collection took place during the months of February-April in Pittsburgh, PA. However, it is interesting to note that some of the data was collected during an unseasonably warm spell in April (80-90° F), but that this warm spell seemed to have little or no discernable effect on subjects' responses to this prompt. In other words, data collected from subjects during this time period (or indeed, at any time) did not mention anything about it being in fact hot in Pennsylvania in April, and thus Jim must have meant *if we want it to be warm and possibly hot when we visit then we should go in April.*

control items when broken down by scalar are noteworthy, as can be seen in Table 12 below:⁵²

Table 12		Target	Control
H&A	<i>some</i>	12.5%	72.9%
	<i>sometimes</i>	8.3%	46.9%
	<i>warm</i>	11.7%	5.2%
H&P	<i>some</i>	55.0%	7.3%
	<i>sometimes</i>	21.7%	3.1%
	<i>warm</i>	34.2%	2.1%

H&A responses are given most often in response to the control items with *some*, significantly less often in the control items with *sometimes* (possibly explained at least in part by a tendency for some subjects to interpret *sometimes* as having a literal reading similar to *occasionally* or *frequently*, as we will discuss in more detail in Section II), and much less often in the control items with *warm*.⁵³ What this result shows is that while subjects did reference higher items on the scale (**H**) in response to the controls containing *warm*, they did not do so in the context of considering alternate utterances the speaker might have said. In fact, subjects only very rarely referred to alternate utterances the speaker might have said (**A**) in response to the *warm* controls (5.2%); whereas this response was very common in reference to the controls containing *some* (74.0%) and to a lesser but still robust degree in the controls containing *sometimes* (47.9%).⁵⁴ We see basically the same story when we compare the **H&A** responses for the control items across the different scalar terms. Thus, we observe that subjects do not seem to arrive at their interpretations in the same way in response to the control prompts containing the scalar term *warm* as they do in response to the control prompts containing *some* and *sometimes*. To sum things up, subjects' responses to the control prompts containing *warm* differed from their responses to the control prompts containing *some* and *sometimes* in two important respects. In the control prompts containing *some* (and to a lesser degree in the control prompts containing *sometimes*) subjects robustly exhibited the standard scalar reasoning pattern (speaker did not say the higher item on the scale, so the speaker must have meant that the higher item on the scale is not a possibility, and thus the subject attributes a strengthened interpretation to the speaker's utterance). This reasoning pattern is virtually absent in subjects' responses to the control prompts containing *warm*. In addition, subjects frequently referred to "lower" items on the scale in response to the control prompts containing the scalar term *warm*, while doing so rarely or never in response to the control items containing the scalar terms *some* or *sometimes*.

Turning to the target items, we examine the **H&P** responses in more detail. As you recall, **H&P** responses were the most frequent reasoning pattern for the target items, in a

⁵² For details of this analysis see Analysis 3.11 in Appendix I.

⁵³ A pairwise comparison of the control prompts shows statistically significant differences between the frequencies of **H&A** responses in control prompts containing *some* vs. *sometimes*, *some* vs. *warm*, and *sometimes* vs. *warm* ($p < .001$). For details of this analysis see Analysis 3.15 in Appendix I.

⁵⁴ For details of the analysis see Analysis 3.11 in Appendix I.

way that closely paralleled the frequency of the **H&A** reasoning pattern in the control items. Examining the **H&P** responses closely, we also see variation in the frequency of **H&P** responses to the different scalar terms within these items, just as we did when we closely examined the **H&A** responses. However, this variation in the **H&P** responses follows a somewhat different pattern than the corresponding pattern in terms of the **H&A** responses when broken down by scalar item. Importantly, while we still see statistically significant differences between the frequencies of **H&P** responses to the target prompts across the different scalar items,⁵⁵ and while the target prompts with *some* are still the most robust when compared with the other scalar items, this difference is much less marked than it was in terms of the **H&A** responses to the control prompts. While **H&A** responses were rarely given in response to the control prompts containing *warm*, **H&P** responses are given almost 35% of the time in response to the target prompts containing *warm*. In fact, if we break down the target prompts containing *warm* question by question, we find that some of these prompts in fact received **H&P** response more on par with the average frequency of **H&P** responses to the target prompts containing *some*, as can be seen in Table 13 below:⁵⁶

Table 13	Q7	Q25	Q29	Q30	Q36
H&P	0.0%	29.2%	54.2%	66.7%	20.8%

Thus, while in one case (Q7) it seems like we have a similar story to tell about the **H&P** responses to the target items containing *warm* as we did in the case of the **H&A** responses to the control items containing *warm* (which is that they rarely (or never) occur), we see that this is in general not the case. Excluding Q7 from the analysis as an outlier, we find that **H&P** responses occur 42.7% of the time,⁵⁷ and the statistically significant difference between the target prompts containing *some* and the target prompts containing *warm* disappears.⁵⁸ Thus, unlike in the control prompts, the reasoning process subjects exhibit in response to the target prompts containing *warm* does not seem to be radically different from the reasoning process exhibited in the other cases, and, in particular, is indistinguishable in most cases from the pattern exhibited by the most robust cases, the target items containing *some*. This result underscores the observation made earlier in discussing the **L** responses, which is that context or domain specific information has a strong effect on the target items across the board and on prompts containing the scalar term *warm* in general, unlike the control prompts with *some* and *sometimes* which are fairly impervious to context or domain specific information. As we saw in Section III of Chapter III, references to context or domain specific information that the subject feels to be relevant to determining whether or not a consequent will still follow if the higher alternative on

⁵⁵ A pairwise comparison of the target prompts shows statistically significant differences between the frequencies of **H&P** responses in target prompts containing *some* vs. *sometimes* ($p < .001$), *some* vs. *warm* ($p < .05$), and *sometimes* vs. *warm* ($p < .05$). For details of this analysis see Analysis 3.15 in Appendix I.

⁵⁶ For details of this analysis see Analysis 3.16 in Appendix I.

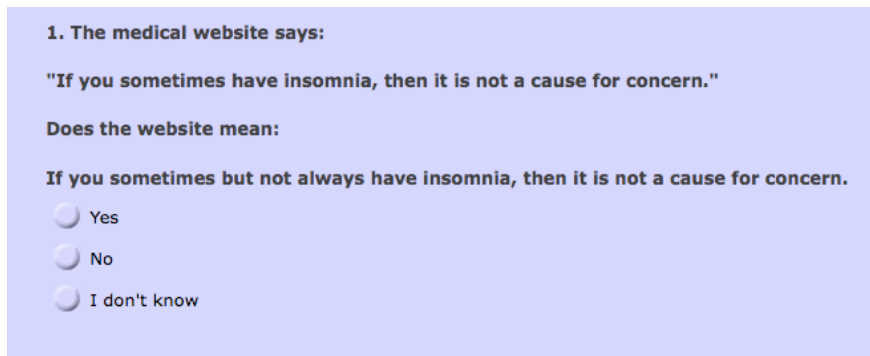
⁵⁷ For details of this analysis see Analysis 3.17 in Appendix I.

⁵⁸ For details of this analysis see Analysis 2.19 in Appendix I.

the scale were the case rather than the lower alternative (e.g. if *all* of the students fail the exam rather than just *some*) is a general feature of the **H&P** responses subjects gave in response to the target items. In conclusion, the main difference that we found between *warm* and *some* or *sometimes* is that *warm* is much more sensitive to context, although obviously *some* and *sometimes* can be affected by it as well (as we saw in the case of the target prompts). More work needs to be done in order to determine whether this is simply an effect of the particular prompts and contexts that we choose,⁵⁹ or whether this is in fact a general fact about *warm*.

Section II: *Sometimes = Occasionally?*

Returning to the observation made earlier in the previous section that *sometimes* is perhaps not a well-behaved scalar item, we consider some of the evidence for this claim. In the previous section we hypothesized that some of the differences in behavior between items containing the scalar items *some* and *sometimes* could be explained by the tendency on the part of subjects to interpret *sometimes* as having a literal meaning similar to *occasionally*. Some examples of this tendency can be seen in some of the responses to Q1, one of the target prompts containing the scalar item *sometimes*:



1. The medical website says:
"If you sometimes have insomnia, then it is not a cause for concern."
Does the website mean:
If you sometimes but not always have insomnia, then it is not a cause for concern.
 Yes
 No
 I don't know

- Subject 102: it's implied that *sometimes* means not always. (Investigator: Because? ... I'm sorry, I'm still just trying to understand why you – what reason you think that's implied.) Subject: Um, because *sometimes* means not all the time? (survey response = yes)
- Subject 115: Um, yes, I would say so. (Investigator: Because?) Because it's just stated in the first clause *if you sometimes have insomnia*, which um, like, *sometimes* is - means *not always*, so *sometimes* ... uh, yeah it's basically the same thing [*sometimes* and *sometimes but not always*]. (survey response = yes)
- Subject 117: Well, I always think *sometimes* – *sometimes* isn't *always* ... so, I would have to say, yes, the website means if you sometimes but not always have insomnia. (Investigator: Could you go over that again, just to make sure I'm clear?)

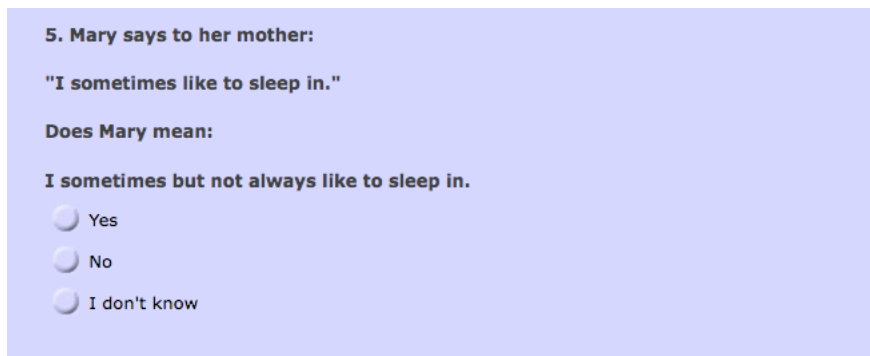
⁵⁹ As we noted, many of our examples had something to do with the weather being warm, or with keeping warm during cold weather. Further studies investigating the question of the scalar status of *warm* would want to examine its behavior in a variety of carefully selected contexts, to see if the context sensitivity that we have observed is uniform across different contexts.

Chapter IV: Further Results: Behavior of Scalar Terms is not Uniform

... about what your reason is?) Okay. Um, when I see the word *sometimes*, *sometimes* means *not always*. I'm sorry ... I feel like I'm just saying what it says on the screen, but ... *sometimes* tends to mean *intermittently* – not steady, not chronic. So, if you're not chronically unable to sleep, then it's not a concern. (survey response = yes)

- Subject 119: I picked "Yes," because ... the website says *sometimes*, and *sometimes* does not mean *always*, which is what that part [the second sentence?] says, so I picked "Yes." (survey response = yes)
- Subject 121: Uh, yes, I would agree with that, because um, they state the word *sometimes*, which means *not always*, and that would mean that it is not a concern. (survey response = yes)
- Subject 126: Yeah, I think so. Um, because, like, *sometimes* and *sometimes but not always* – *sometimes* basically means *but not always*, so there's a repetition of the same word. (survey response = yes)

This type of response occurred in the control prompts that contained the scalar item *sometimes* as well, as can be seen below in the sample responses to Q5:



5. Mary says to her mother:
"I sometimes like to sleep in."
Does Mary mean:
I sometimes but not always like to sleep in.
 Yes
 No
 I don't know

- Subject 117: This one is kind of more difficult, because rather than just looking at the word *sometimes*, I'm also looking at the "like to" part, because it could be that she's saying she doesn't *actually* like to sleep in except for sometimes, or it could be that she's saying she doesn't *actually get* to sleep in except for sometimes. I'm fairly certain that *sometimes* means *not always*, but ... I'm not quite sure about whether she is referring to actually getting to sleep in, or if she actually likes sleeping in ... I'm going to have to go with "I don't know" on this one. (survey response = don't know)
- Subject 119: I picked "Yes" again, meaning *I sometimes but not always like to sleep in* is what she [Mary] is saying, because she says *I sometimes like to sleep in*, and that means not all the time, so it's only sometimes. (survey response = yes)
- Subject 126: I find it's the same reasoning as the previous question [Question 4] – that *sometimes* has the same meaning as *not always* – you could view them as synonyms, thus the answer would be yes, she means the same [thing in both sentences]. (survey response = yes)

Thus, it seems that *sometimes* is not simply a compound form of *some time*, at least for some people in at least some linguistic contexts, and thus it does not behave exactly like

some. However, as we can see from these early examples (Q1 and Q5), the number of people who react to *sometimes* in this way is actually relatively small.⁶⁰ Additionally, it is interesting to note that although these subjects initially reacted to *sometimes* this way, very few stuck with this interpretation consistently,⁶¹ and most seemed willing to abandon this "literal meaning" reading of *sometimes* as *occasionally* or *infrequently* in the later prompts. For example, consider some of the responses to the following prompt:

4. On the first day of class, the professor says:
"If students are sometimes late for class, then it makes me upset."
Does the professor mean:
If students are sometimes but not always late for class, then it makes me upset.
 Yes
 No
 I don't know

- Subject 102: Okay, well ... you can't really know – if I was the professor, I wouldn't really know if the *same* student was late like, consistently, or not, so I would assume That He means like, in general, that when students are late, then it makes him upset, but it doesn't really mean like, sometimes and not always. I feel like the sentence – or the meaning – the implication of his statement is, like, trying to say that you shouldn't be late for class, even though I [the professor] might not know if you're always, or not always, late. (survey response = no)
- Subject 121: Um, yes, I'm sure he probably would not like if anyone is ever late, even if it is on occasion – he might let it slide, but I'm sure he would still be upset. (survey response = yes)

These subjects are among those who used the "literal meaning" interpretation of *sometimes* in determining their answer to Q1, but who never used it again. In general, most of the people who began with the "literal meaning" interpretation of *sometimes* quickly moved away from this type of interpretation after the first few examples. By the fourth prompt involving *sometimes* (Q10), all of the subjects who started out with this "literal meaning" interpretation in response to the first prompt had abandoned this approach, with two exceptions. The first exception is Subject 126, who stuck with the "literal meaning" interpretation of *sometimes* throughout the experiment, despite changes of context. The other exception is Subject 119, who used the "literal meaning" interpretation consistently in response to the control prompts containing *sometimes*, but not explicitly in response to all of the target prompts containing *sometimes*, although it is

⁶⁰ Although this "literal meaning" interpretation of *sometimes* was not initially coded for, a second run-through of the data indicates that this interpretation of *sometimes* seems to be confined to six subjects (102, 112, 117, 119, 121, 126), all of who are represented in the first example above.

⁶¹ There were a couple of exceptions, which will be discussed in more detail shortly.

possible that the subject is doing so implicitly. This subject consistently strengthened all of the prompts containing *sometimes*. The same is also true for all of the subjects that we looked at in our first example, except for subject 102, who gave a strengthened interpretation to ~50% of the prompts containing *sometimes*. Thus, it is hard to say precisely whether or not these 5 subjects used this "literal meaning" interpretation of *sometimes* throughout the experiment. Regardless, it seems clear that whatever effect this tendency to interpret *sometimes* as *occasionally* (and thus *not always*) may have had on the experimental outcomes, it was a fairly marginal effect, and thus is at best only a partial answer to the difference between the behavior of the scalar terms *sometimes* and *some*. Perhaps a better answer to the difference between *some* and *sometimes* is that subjects often seemed to find the prompts containing *sometimes* to be oddly phrased or unnatural, particularly in terms of the target prompts. This seems like it could possibly account for at least some of the rest of the differences between these two otherwise quite similar scalars, and thus the differences between *sometimes* and *some* does not seem as striking as the differences we observed between *some* and *warm*.

Section III: Further Research

Some interesting possibilities for further research are suggested by all of the results from this experiment. First, we'd like to know more about the **H&P** reasoning pattern that emerged from the subjects' think-aloud responses to the target prompts. Although the results were not as robust as one might like (only occurred in 39.6% of the target items), they are as robust as the **H&A** reasoning pattern in the control prompts (41.7%). One possibility that was suggested at the end of Chapter III was that subjects needed to "reach" more to arrive at an unstrengthened reading and thus were more likely to feel the need to "walk through" the reasoning process explicitly, thus increasing the rate of **H&P** responses for those items. A further experiment that presented subjects with some target prompts (conditionals) that gave a strengthened interpretation as the potential meaning of the utterance (e.g., *some but not all*) and some that gave an unstrengthened interpretation as the potential meaning (e.g., *some and possibly all*) might have the effect of bringing out this **H&P** response pattern more explicitly. Also, it might be interesting to examine different types of conditionals with embedded scalar terms and see if subjects react differently to different types of conditionals (e.g., epistemic vs. counter-factual conditionals). Based on the results of this study, the recommendation would be to focus on the scalar terms *some* and *sometimes* as the embedded scalars. The atypical behavior of *warm* muddied the waters somewhat in terms of the analysis of the conditional statements, although producing interesting results of their own.

In fact, the results about *warm* are as potentially significant as the results that we obtained about scalar terms in the antecedent of conditionals. We did not set out to investigate the different scalar terms, but it turns out that there are significant differences between them. This leads one to wonder about other Horn scales (e.g. *<excellent, good>* or *<beautiful, pretty>*) and whether they also might display behavior similar to the *<hot, warm>* scale. Our research in this experiment suggests that the behavior of the different Horn scales is not uniform, and that perhaps some of them are not really scales at all, or that they are scales of a very different sort.

Chapter IV: Further Results: Behavior of Scalar Terms is not Uniform

These two different areas of interest could be combined into one experiment. The prompts for the experiment could be split into three types:

- Target Prompts 1: Scalar terms *some* and *sometimes* embedded under in the antecedent of conditionals, in order to further investigate the way people reason about these types of utterances
- Target Prompts 2: Investigation of various Horn scales (including *warm*) in unembedded contexts, to investigate the behavior of more Horn scales
- Control Prompts: Unembedded scalar implicatures involving *some*

Ideally, target prompts 1 and 2 could serve as the distractor for each other, with the unembedded scalar implicatures with *some* serving as the control for both types of target prompts. The prompts could be designed along similar lines to the prompts used in this experiment, and a think-aloud protocol should be used again in order to "look under the hood" and see what is going on when people try to reason about these various linguistic phenomenon.

References

- Carston, R. (1988). Implicature, explicature and truth-theoretic semantics. In R. M. Kempson (Ed.), *Mental representations : The interface between language and reality* (pp. 155-182). Cambridge ; New York: Cambridge University Press.
- Chierchia, G. (2004). Scalar implicatures, polarity phenomena and the Syntax/Pragmatics interface. In A. Belletti (Ed.), *Structures and beyond* (pp. 39-103). Oxford ; New York: Oxford University Press.
- Chierchia, G. (2006). Broaden your views: Implicatures of domain widening and the "logicality" of language. *Linguistic Inquiry*, 37(4), pp. 535-590.
- Cohen, L. J. (1971). The logical particles of natural language. In Y. Bar-Hillel (Ed.), *Pragmatics of natural language* (pp. 50-68). Dordrecht: Reidel.
- Fox, D. (2007). Free choice and the theory of scalar implicatures. In U. Sauerland, & P. Stateva (Eds.), *Presupposition and implicature in compositional semantics* (pp. 71-120). Basingstoke ; New York: Palgrave Macmillan.
- Gazdar, G. (1979). *Pragmatics : Implicature, presupposition and logical form*. New York: Academic Press.
- Geurts, B., & Pouscoulous, N. (2009). Embedded implicatures?!? *Semantics & Pragmatics*, 2, 9/14/2009. doi:10.3765/sp.2.4
- Geurts, B. (2009). Scalar implicature and local pragmatics. *Mind & Language*, 24(1), 51-79. doi:10.1111/j.1468-0017.2008.01353.x
- Grice, H. P. (1989). Logic and conversation. In *Studies in the way of words* (pp. 22-40). Cambridge, Mass: Harvard University Press.
- Horn, L. R. (1984). Toward a new taxonomy for pragmatic inference: Q-based and R-based implicature. In D. Schiffrin (Ed.), *Meaning, form, and use in context: Linguistic applications* (pp. 11-42). Washington, D.C.: Georgetown University Press.
- Landman, F. (1998). Plurals and maximalization. In S. D. Rothstein (Ed.), *Events and grammar* (pp. 237-271). Dordrecht : Boston: Kluwer Academic.
- Landman, F. (2000). *Events and plurality : The jerusalem lectures*. Dordrecht ; Boston: Kluwer Academic Publishers.
- Levinson, S. C. (2000). *Presumptive meanings : The theory of generalized conversational implicature*. Cambridge, Mass: MIT Press.
- Recanati, F. (2003). Embedded implicatures. *Nous-Supplement: Philosophical Perspectives*, 17, 299-332.

References

Sauerland, U. (2004). Scalar implicatures in complex sentences. *Linguistics and Philosophy*, 27(3), 367-391.

Simons, M. (2010). A gricean view on intrusive implicatures. Preprint manuscript available at: [http://www.hss.cmu.edu/philosophy/simons/Gricean view on Intrusive Implicatures.pdf](http://www.hss.cmu.edu/philosophy/simons/Gricean%20view%20on%20Intrusive%20Implicatures.pdf). Published in K. Petrus (Ed.), *Meaning and analysis : New essays on grice* (pp. 138-169). Basingstoke: Palgrave Macmillan.

Part I: Analysis of Coder Agreement

Data file: coderagreement062710.sav
Number of subjects: 3 (101, 108, 109)
Coder 1: Primary coder
Coder 2: Outside coder

Analysis 1.1

Questions:

1. What is the coder agreement for the **H** code?
2. What is the coder agreement for the **L** code?
3. What is the coder agreement for the **A** code?
4. What is the coder agreement for the **P** code?
5. What is the coder agreement for the **CS** code?

Analysis 1.1	Kappa
1. H1 x H2	.593
2. L1 x L2	.937
3. A1 x A2	.778
4. P1 x P2	.634
5. CS1 x CS2	.509

Part II: Analysis of Survey Responses

Data file: data062710_1.sav
Number of subjects: 26 (10 male, 16 female)
Number of target prompts per subject: 15
Number of control prompts per subject: 12
Number of distractor prompts per subject: 12 (removed from analysis)

Analysis 2.1

Questions:

1. Conditional on $qtype=target$, what is the frequency of Yes/No/I don't know responses?
2. Conditional $qtype=control$, what is the frequency of Yes/No/I don't know responses?

Analysis 2.1	Yes	No	I don't know
1. Target	52.6%	42.3%	5.1%
2. Control	74.4%	22.4%	3.2%

Analysis 2.2

Questions:

1. Conditional on ($qtype=target$ & $scalar=some$), what is the frequency of Yes/No/I don't know responses?

Appendix I: Data Analysis

2. Conditional on (qtype=target & scalar=*sometimes*), what is the frequency of Yes/No/I don't know responses?
3. Conditional on (qtype=target & scalar=*warm*), what is the frequency of Yes/No/I don't know responses?
4. Conditional on (qtype=control & scalar=*some*), what is the frequency of Yes/No/I don't know responses?
5. Conditional on (qtype=control & scalar=*sometimes*), what is the frequency of Yes/No/I don't know responses?
6. Conditional on (qtype=control & scalar=*warm*), what is the frequency of Yes/No/I don't know responses?

Analysis 2.2		Yes	No	I don't know
Target	1. <i>Some</i>	50.8%	45.4%	3.8%
	2. <i>Sometimes</i>	64.6%	33.1%	2.3%
	3. <i>Warm</i>	42.3%	48.5%	9.2%
Control	4. <i>Some</i>	96.2%	1.9%	1.9%
	5. <i>Sometimes</i>	91.3%	6.7%	1.9%
	6. <i>Warm</i>	35.6%	58.7%	5.8%

Analysis 2.3

Questions:

1. Conditional on (scalar=*some* v scalar=*sometimes*), is the difference in the distribution of the Yes/No/I don't know responses in the target vs. control items statistically significant?

Analysis 2.3	Chi Square	<i>df</i>	<i>p</i>	Conclusion
1. QTYPE x RESPONSE	80.336	2	.000	Association $p < .001$

Analysis 2.4

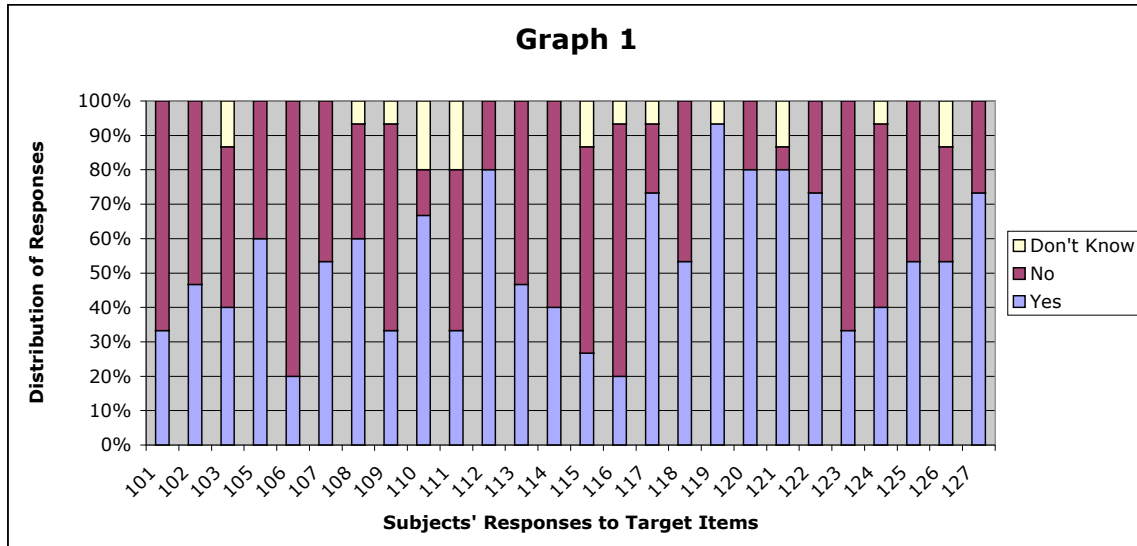
Questions:

1. For each subject, what is the frequency of Yes/No/I don't know answers, conditional on qtype=target?
2. For each subject, what is the frequency of Yes/No/I don't know answers, conditional on qtype=control?

Analysis 2.4		1. Target: Yes/ No/ I don't know	2. Control: Yes/ No/ I don't know
ID	Gender		
101	M	33.3/ 66.7/ 0.0	66.7/ 33.3/ 0.0
102	F	46.7/ 53.3/ 0.0	41.7/ 50.0/ 8.3
103	F	40.0/ 46.7/ 13.3	58.3/ 25.0/ 16.7
105	M	60.0/ 40.0/ 0.0	100.0/ 0.0/ 0.0
106	M	20.0/ 80.0/ 0.0	50.0/ 50.0/ 0.0
107	M	53.3/ 46.7/ 0.0	83.3/ 16.7/ 0.0

Appendix I: Data Analysis

108	M	60.0/ 33.3/ 6.7	66.7/ 33.3/ 0.0
109	F	33.3/ 60.0/ 6.7	58.3/ 33.3/ 8.3
110	M	66.7/ 13.3/ 20.0	100.0/ 0.0/ 0.0
111	F	33.3/ 46.7/ 20.0	67.7/ 33.3/ 0.0
112	F	80.0/ 20.0/ 0.0	100.0/ 0.0/ 0.0
113	M	46.7/ 53.3/ 0.0	67.7/ 33.3/ 0.0
114	M	40.0/ 60.0/ 0.0	67.7/ 33.3/ 0.0
115	F	26.7/ 60.0/ 13.3	58.3/ 25.0/ 16.7
116	M	20.0/ 73.3/ 6.7	66.7/ 33.3/ 0.0
117	F	73.3/ 20.0/ 6.7	91.7/ 0.0/ 8.3
118	F	53.3/ 46.7/ 0.0	75.0/ 16.7/ 8.3
119	F	93.3/ 0.0/ 6.7	100.0/ 0.0/ 0.0
120	F	80.0/ 20.0/ 0.0	100.0/ 0.0/ 0.0
121	F	80.0/ 6.7/ 13.3	100.0/ 0.0/ 0.0
122	M	73.3/ 26.7/ 0.0	83.3/ 16.7/ 0.0
123	F	33.3/ 66.7/ 0.0	50.0/ 50.0/ 0.0
124	F	40.0/ 53.3/ 6.7	66.7/ 33.3/ 0.0
125	F	53.3/ 46.7/ 0.0	66.7/ 33.3/ 0.0
126	F	53.3/ 33.3/ 13.3	75.0/ 8.3/ 16.7
127	F	73.3/ 26.7/ 0.0	75.0/ 25.0/ 0.0



Analysis 2.5

Questions:

1. Conditional on qtype=target, is the difference in the distribution of Yes/No/I don't know responses in males vs. females statistically significant?
2. Conditional on qtype=control, is the difference in the distribution of Yes/No/I don't know responses in males vs. females statistically significant?

Appendix I: Data Analysis

Analysis 2.5	Chi Square	<i>df</i>	<i>p</i>	Conclusion
RESPONSE x GENDER				
1. Target	3.735	2	.155	Independent
2. Control	6.9811	2	.030	Association $p < .05$

Analysis 2.6

Questions:

1. Conditional on *qtype*=control, what is the frequency of Yes/No/I don't know responses among the male subjects?
2. Conditional on *qtype*=control, what is the frequency of Yes/No/I don't know responses among the female subjects?

Analysis 2.6	Yes	No	I don't know
1. Men (<i>n</i> =10)	75.0%	25.0%	0.0%
2. Women (<i>n</i> =16)	74.0%	20.8%	5.2%

Analysis 2.7

Questions:

1. Is the difference in the distribution of Yes/No responses in the strengthened vs. "truth-functionally" strengthened prompts statistically significant?
2. Is the difference in the distribution of Yes/No responses in the strengthened vs. unstrengthened prompts statistically significant?
3. Is the difference in the distribution of Yes/No responses in the strengthened vs. neutral prompts statistically significant?
4. Is the difference in the distribution of Yes/No responses in the "truth-functionally" strengthened vs. unstrengthened prompts statistically significant?
5. Is the difference in the distribution of Yes/No responses in the "truth-functionally" strengthened vs. neutral prompts statistically significant?
6. Is the difference in the distribution of Yes/No responses in the unstrengthened vs. neutral prompts statistically significant?

Notes:

1. "I don't know" responses removed from this analysis as outliers

Appendix I: Data Analysis

Analysis 2.7	Chi Square	<i>df</i>	<i>p</i>	Conclusion
1. Strengthened vs. TF Strengthened	8.500	1	.004	Association $p < .05$
2. Strengthened vs. Unstrengthened	24.303	1	.000	Association $p < .001$
3. Strengthened vs. Neutral	19.229	1	.000	Association $p < .001$
4. TF Strengthened vs. Unstrengthened	66.253	1	.000	Association $p < .001$
5. TF Strengthened vs. Neutral	59.794	1	.000	Association $p < .001$
6. Unstrengthened vs. Neutral	1.090	1	.296	Independent

Analysis 2.8

Questions:

1. Is the difference between the frequency of Yes vs. No answers in the 6 neutral prompts a statistically significant difference?
2. Is the difference between the frequency of Yes vs. No answers in the 2 strengthened prompts a statistically significant difference?
3. Is the difference between the frequency of Yes vs. No answers in the 3 "truth-functionally" strengthened prompts a statistically significant difference?
4. Is the difference between the frequency of Yes vs. No answers in the 4 unstrengthened prompts a statistically significant difference?

Notes:

1. "I don't know" responses removed from this analysis as outliers
2. The analysis of the "truth-functionally" strengthened prompts fails to meet the assumptions for Chi-square. 3 cells (50.0%) have expected count less than 5. The minimum expected count is 1.30. However, this is not really a concern, since the results show these prompts to be independent, and thus there is no need to worry that the low expected counts are causing the prompts to appear to be associated when they really are not.

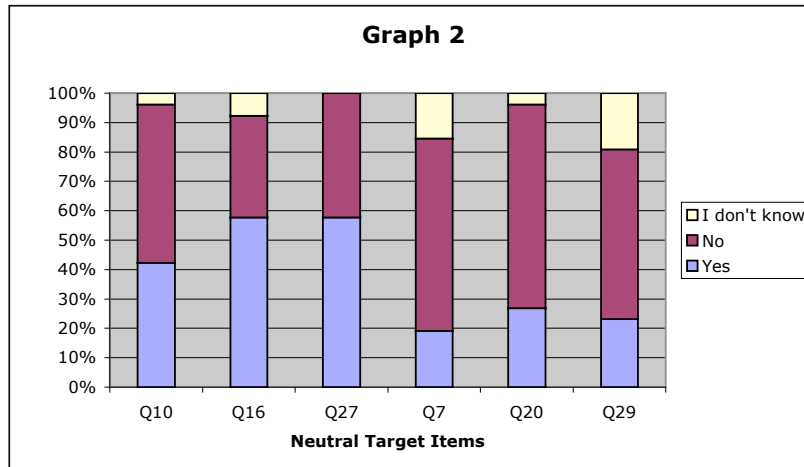
Analysis 2.8	Chi Square	<i>df</i>	<i>p</i>	Conclusion
1. Neutral	13.768	5	.017	Association $p < .05$
2. Strengthened	3.787	1	.052	Independent
3. TF Strengthened	.591	2	.744	Independent
4. Unstrengthened	6.565	3	.087	Independent

Analysis 2.9

Questions:

1. What is the frequency of Yes/No/I don't know responses for the neutral prompt Q7?
2. What is the frequency of Yes/No/I don't know responses for the neutral prompt Q10?
3. What is the frequency of Yes/No/I don't know responses for the neutral prompt Q16?
4. What is the frequency of Yes/No/I don't know responses for the neutral prompt Q20?
5. What is the frequency of Yes/No/I don't know responses for the neutral prompt Q27?
6. What is the frequency of Yes/No/I don't know responses for the neutral prompt Q29?

Analysis 2.9	Yes	No	I don't know
1. Q7	19.2%	65.4%	15.4%
2. Q10	42.3%	53.8%	3.8%
3. Q16	57.7%	34.6%	7.7%
4. Q20	26.9%	69.2%	3.8%
5. Q27	57.7%	42.3%	0.0%
6. Q29	23.1%	57.7%	19.2%



Analysis 2.10

Questions:

1. Is the difference between the frequency of Yes vs. No answers to the neutral prompts Q10, Q16, Q27 a statistically significant difference?
2. Is the difference between the frequency of Yes vs. No answers to the neutral prompts Q7, Q20, Q29 a statistically significant difference?
3. Is the difference between the cluster of neutral items (Q7, Q20, Q29) and the unstrengthened items a statistically significant difference?

Appendix I: Data Analysis

Notes:

- "I don't know" responses removed from this analysis as outliers

Analysis 2.10	Chi Square	<i>df</i>	<i>p</i>	Conclusion
1. Q10 vs. Q16 vs. Q27	1.838	2	.399	Independent
2. Q7 vs. Q20 vs. Q27	.236	2	.889	Independent
3. (Q10, Q16, Q27) vs. Unstrengthened prompts	8.384	6	.211	Independent

Analysis 2.11

Questions:

- What is the frequency of Yes/No/I don't know responses for the *sometimes* target prompt Q1?
- What is the frequency of Yes/No/I don't know responses for the *sometimes* target prompt Q14?
- What is the frequency of Yes/No/I don't know responses for the *sometimes* target prompt Q4?
- What is the frequency of Yes/No/I don't know responses for the *sometimes* target prompt Q10?
- What is the frequency of Yes/No/I don't know responses for the *sometimes* target prompt Q27?

Analysis 2.11		Yes	No	I don't know
Category	Question #			
S	1. Q1	91.7%	8.3%	0.0%
TF	2. Q14	91.7%	4.2%	4.2%
U	3. Q4	45.8%	50.0%	4.2%
N	4. Q10	37.5%	58.3%	4.2%
	5. Q27	54.2%	45.8%	0.0%

Analysis 2.12

Questions:

- Is the difference between the frequency of Yes vs. No answers to the strengthened target prompts Q1 and Q36 a statistically significant difference?

Notes:

- "I don't know" responses removed from this analysis as outliers

Analysis 2.12	Chi Square	<i>df</i>	<i>p</i>	Conclusion
1. Q1 vs. Q36	5.301	1	.021	Association $p < .05$

Analysis 2.13

Questions:

- What is the frequency of Yes/No/I don't know responses for the strengthened *warm* target prompt Q36?

Appendix I: Data Analysis

Analysis 2.13	Yes	No	I don't know
1. Q36	58.3%	33.3%	8.3%

Analysis 2.14

Questions:

1. What is the frequency of Yes/No/I don't know responses for the *warm* control prompt Q9?
2. What is the frequency of Yes/No/I don't know responses for the *warm* control prompt Q19?
3. What is the frequency of Yes/No/I don't know responses for the *warm* control prompt Q28?
4. What is the frequency of Yes/No/I don't know responses for the *warm* control prompt Q33?

Analysis 2.14	Yes	No	I don't know
1. Q9	34.6%	57.7%	7.7%
2. Q19	30.8%	65.4%	3.8%
3. Q28	42.3%	53.8%	3.8%
4. Q33	34.7%	57.7%	7.7%

Part III: Analysis of Think-Aloud Responses

Data file: data062710_2.sav

Number of subjects: 24 (9 male, 15 female)

- Data from subjects 103 and 107 were removed from the analysis of the think-aloud responses because half of the think-aloud data for these subjects was lost due to equipment malfunction.

Number of target prompts per subject: 15

Number of control prompts per subject: 12

Number of distractor prompts per subject: 12 (removed from analysis)

Analysis 3.1

Questions:

1. How frequently do **H** responses occur in response to the prompts?
2. How frequently do **L** responses occur in response to the prompts?
3. How frequently do **A** responses occur in response to the prompts?
4. How frequently do **P** responses occur in response to the prompts?
5. How frequently do **CS** responses occur in response to the prompts?

Analysis 3.1	1. H	2. L	3. A	4. P	5. CS
	63.4%	15.7%	25.8%	24.7%	25.6%

Analysis 3.2

Questions:

1. Within the target prompts:
 - a. How frequently do **H** responses occur in response to the prompts?
 - b. How frequently do **L** responses occur in response to the prompts?
 - c. How frequently do **A** responses occur in response to the prompts?
 - d. How frequently do **P** responses occur in response to the prompts?
 - e. How frequently do **CS** responses occur in response to the prompts?
 - f. How frequently do **H&A** responses occur in response to the prompts?
 - g. How frequently do **H&P** responses occur in response to the prompts?
2. Within the control prompts:
 - a. How frequently do **H** responses occur in response to the prompts?
 - b. How frequently do **L** responses occur in response to the prompts?
 - c. How frequently do **A** responses occur in response to the prompts?
 - d. How frequently do **P** responses occur in response to the prompts?
 - e. How frequently do **CS** responses occur in response to the prompts?
 - f. How frequently do **H&A** responses occur in response to the prompts?
 - g. How frequently do **H&P** responses occur in response to the prompts?

Analysis 3.2	1. Target	2. Control
a. H	61.7% (<i>n</i> =222)	65.6% (<i>n</i> =189)
b. L	15.6% (<i>n</i> =56)	16.0% (<i>n</i> =46)
c. A	12.5% (<i>n</i> =45)	42.4% (<i>n</i> =122)
d. P	40.3% (<i>n</i> =145)	5.2% (<i>n</i> =15)
e. CS	26.9% (<i>n</i> =97)	24.0% (<i>n</i> =69)
f. H&A	10.8% (<i>n</i> =39)	41.7% (<i>n</i> =120)
g. H&P	36.9% (<i>n</i> =133)	4.2% (<i>n</i> =12)

Analysis 3.3

Questions:

1. Is the difference between the frequency of the occurrence of **H** responses in the control vs. target prompts statistically significant?
2. Is the difference between the frequency of the occurrence of **L** responses in the control vs. target prompts statistically significant?
3. Is the difference between the frequency of the occurrence of **A** responses in the control vs. target prompts statistically significant?
4. Is the difference between the frequency of the occurrence of **P** responses in the control vs. target prompts statistically significant?
5. Is the difference between the frequency of the occurrence of **CS** responses in the control vs. target prompts statistically significant?
6. Is the difference between the frequency of the occurrence of **H&A** responses in the control vs. target prompts statistically significant?
7. Is the difference between the frequency of the occurrence of **H&P** responses in the control vs. target prompts statistically significant?

Appendix I: Data Analysis

Analysis 3.3	Chi-Square Value	<i>df</i>	<i>p</i>	Conclusion
1. H x QTYPE	1.081	1	.299	Independent
2. L x QTYPE	.021	1	.885	Independent
3. A x QTYPE	74.580	1	.000	Association $p < .001$
4. P x QTYPE	105.825	1	.000	Association $p < .001$
5. CS x QTYPE	.749	1	.387	Independent
6. H&A x QTYPE	82.150	1	.000	Association $p < .001$
7. H&P x QTYPE	98.968	1	.000	Association $p < .001$

Analysis 3.4

Questions:

1. Conditional on **P**=yes, what is the frequency of **H** responses to the target prompts?
2. Conditional on **A**=yes, what is the frequency of **H** responses to the control prompts?

Analysis 3.4	% of prompts that received H responses
1. Target (P =yes)	93.1%
2. Control (A = yes)	99.2%

Analysis 3.5

Questions:

1. Conditional on (qtype=target & response=yes): ($n=191$)
 - a. How frequently do **H** responses occur in response to the prompts?
 - b. How frequently do **L** responses occur in response to the prompts?
 - c. How frequently do **A** responses occur in response to the prompts?
 - d. How frequently do **P** responses occur in response to the prompts?
 - e. How frequently do **CS** responses occur in response to the prompts?
 - f. How frequently do **H&A** responses occur in response to the prompts?
 - g. How frequently do **H&P** responses occur in response to the prompts?
2. Conditional on (qtype=target & response=no): ($n=151$)
 - a. How frequently do **H** responses occur in response to the prompts?
 - b. How frequently do **L** responses occur in response to the prompts?
 - c. How frequently do **A** responses occur in response to the prompts?
 - d. How frequently do **P** responses occur in response to the prompts?
 - e. How frequently do **CS** responses occur in response to the prompts?
 - f. How frequently do **H&A** responses occur in response to the prompts?
 - g. How frequently do **H&P** responses occur in response to the prompts?
3. Conditional on (qtype=control & response=yes): ($n=215$)
 - a. How frequently do **H** responses occur in response to the prompts?
 - b. How frequently do **L** responses occur in response to the prompts?
 - c. How frequently do **A** responses occur in response to the prompts?
 - d. How frequently do **P** responses occur in response to the prompts?
 - e. How frequently do **CS** responses occur in response to the prompts?

Appendix I: Data Analysis

- f. How frequently do **H&A** responses occur in response to the prompts?
- g. How frequently do **H&P** responses occur in response to the prompts?
- 4. Conditional on (qtype=control & response=no): ($n=65$)
 - a. How frequently do **H** responses occur in response to the prompts?
 - b. How frequently do **L** responses occur in response to the prompts?
 - c. How frequently do **A** responses occur in response to the prompts?
 - d. How frequently do **P** responses occur in response to the prompts?
 - e. How frequently do **CS** responses occur in response to the prompts?
 - f. How frequently do **H&A** responses occur in response to the prompts?
 - g. How frequently do **H&P** responses occur in response to the prompts?

Notes:

- 1. "I don't know" responses removed from this analysis as outliers

Analysis 3.5	Target		Control	
	1. Yes	2. No	3. Yes	4. No
a. H	57.6%	68.2%	78.1%	30.8%
b. L	6.8%	25.2%	3.3%	55.4%
c. A	17.8%	6.6%	56.7%	0.0%
d. P	29.8%	53.6%	5.1%	4.6%
e. CS	28.8%	21.2%	26.0%	13.8%
f. H&A	16.8%	4.0%	55.8%	0.0%
g. H&P	27.2%	49.7%	4.2%	4.6%

Analysis 3.6

Questions:

- 1. Conditional on (**L**=yes & qtype=target), what percentage of the items contain the scalar *warm* as opposed to *some* or *sometimes*?
- 2. Conditional on (**L**=yes & qtype=control), what percentage of the items contain the scalar *warm* as opposed to *some* or *sometimes*?

Notes:

- 1. **L**=yes & qtype=target: $n = 56$
- 2. **L**=yes & qtype=control: $n = 46$

Analysis 3.6	1. Targets			2. Controls		
	<i>Some</i>	<i>Sometimes</i>	<i>Warm</i>	<i>Some</i>	<i>Sometimes</i>	<i>Warm</i>
	19.6%	11.8%	68.6%	9.3%	0.0%	90.7%

Analysis 3.7

Questions:

- 1. Conditional on (qtype=control & response=yes), what is the overlap between **H&A** responses and **L** responses? In other words, how many times did subjects give both a **H&A** response and a **L** response to the same item?
- 2. Conditional on (qtype=control & response=yes), what is the overlap between **H&A** responses and **CS** responses? In other words, how many times did subjects give both a **H&A** response and an **CS** response to the same item?

Appendix I: Data Analysis

3. Conditional on (qtype=control & response=yes), what is the overlap between **H&A** responses and **H&P** responses? In other words, how many times did subjects give both a **H&A** response and a **H&P** response to the same item?

Notes:

1. qtype=control & response=yes: $n = 215$ items

Analysis 3.7	1. H&A x L	2. H&A x CS	3. H&A x H&P
	3.3% ($n=4$)	14.2% ($n=17$)	2.5% ($n=3$)

Analysis 3.8

Questions:

1. Within the target prompts:
 - a. Conditional on **H**=yes, how frequently do subjects respond "yes" vs. "no?"
 - b. Conditional on **L**=yes, how frequently do subjects respond "yes" vs. "no?"
 - c. Conditional on **A**=yes, how frequently do subjects respond "yes" vs. "no?"
 - d. Conditional on **P**=yes, how frequently do subjects respond "yes" vs. "no?"
 - e. Conditional on **CS**=yes, how frequently do subjects respond "yes" vs. "no?"
 - f. Conditional on **H&A**=yes, how frequently do subjects respond "yes" vs. "no?"
 - g. Conditional on **H&P**=yes, how frequently do subjects respond "yes" vs. "no?"
2. Within the control prompts:
 - a. Conditional on **H**=yes, how frequently do subjects respond "yes" vs. "no?"
 - b. Conditional on **L**=yes, how frequently do subjects respond "yes" vs. "no?"
 - c. Conditional on **A**=yes, how frequently do subjects respond "yes" vs. "no?"
 - d. Conditional on **P**=yes, how frequently do subjects respond "yes" vs. "no?"
 - e. Conditional on **CS**=yes, how frequently do subjects respond "yes" vs. "no?"
 - f. Conditional on **H&A**=yes, how frequently do subjects respond "yes" vs. "no?"
 - g. Conditional on **H&P**=yes, how frequently do subjects respond "yes" vs. "no?"

Notes:

1. "I don't know" responses included in the analysis but not reported.
2. See Analysis 11 for the number of items within each of these categories.

Appendix I: Data Analysis

Analysis 3.8	1. Targets		2. Controls	
	Yes	No	Yes	No
a. H	49.5%	46.4%	88.9%	10.6%
b. L	23.2%	67.9%	15.2%	78.3%
c. A	75.6%	22.2%	100.0%	0.0%
d. P	39.3%	55.9%	73.3%	20.0%
e. CS	56.7%	33.3%	81.2%	13.0%
f. H&A	82.1%	15.4%	100.0%	0.0%
g. H&P	39.1%	56.4%	75.0%	25.0%

Analysis 3.9

Questions:

1. Conditional on (qtype=target & **H&A**=yes), how many of these items were in the "neutral" category (**N**) vs. the strengthened (**S**), "truth-functionally" strengthened (**TF**) or unstrengthened (**U**) categories?

Analysis 3.9	N	S	TF	U
1. qtype=target & H&A =yes	<i>n</i> = 18	<i>n</i> = 7	<i>n</i> = 6	<i>n</i> = 8

Analysis 3.10

Questions:

1. Conditional on **L**=yes, what percentage of the prompts contain the scalar *warm* as opposed to *some* or *sometimes*?

Analysis 3.10	<i>warm</i>	<i>some</i>	<i>sometimes</i>
1. L =yes	80.4%	13.7%	5.9%

Analysis 3.11

Questions:

1. Conditional on (qtype=target & scalar=*some*):
 - a. How frequently do subjects give a **H** response?
 - b. How frequently do subjects give a **L** response?
 - c. How frequently do subjects give an **A** response?
 - d. How frequently do subjects give a **P** response?
 - e. How frequently do subjects give a **CS** response?
 - f. How frequently do subjects give a **H&A** response?
 - g. How frequently do subjects give a **H&P** response?
2. Conditional on (qtype=target & scalar=*sometimes*):
 - a. How frequently do subjects give a **H** response?
 - b. How frequently do subjects give a **L** response?
 - c. How frequently do subjects give an **A** response?
 - d. How frequently do subjects give a **P** response?
 - e. How frequently do subjects give a **CS** response?
 - f. How frequently do subjects give a **H&A** response?

Appendix I: Data Analysis

- g. How frequently do subjects give a **H&P** response?
- 3. Conditional on (qtype=target & scalar=*warm*):
 - a. How frequently do subjects give a **H** response?
 - b. How frequently do subjects give a **L** response?
 - c. How frequently do subjects give an **A** response?
 - d. How frequently do subjects give a **P** response?
 - e. How frequently do subjects give a **CS** response?
 - f. How frequently do subjects give a **H&A** response?
 - g. How frequently do subjects give a **H&P** response?
- 4. Conditional on (qtype=control & scalar=*some*):
 - a. How frequently do subjects give a **H** response?
 - b. How frequently do subjects give a **L** response?
 - c. How frequently do subjects give an **A** response?
 - d. How frequently do subjects give a **P** response?
 - e. How frequently do subjects give a **CS** response?
 - f. How frequently do subjects give a **H&A** response?
 - g. How frequently do subjects give a **H&P** response?
- 5. Conditional on (qtype=control & scalar=*sometimes*):
 - a. How frequently do subjects give a **H** response?
 - b. How frequently do subjects give a **L** response?
 - c. How frequently do subjects give an **A** response?
 - d. How frequently do subjects give a **P** response?
 - e. How frequently do subjects give a **CS** response?
 - f. How frequently do subjects give a **H&A** response?
 - g. How frequently do subjects give a **H&P** response?
- 6. Conditional on (qtype=control & scalar=*warm*):
 - a. How frequently do subjects give a **H** response?
 - b. How frequently do subjects give a **L** response?
 - c. How frequently do subjects give an **A** response?
 - d. How frequently do subjects give a **P** response?
 - e. How frequently do subjects give a **CS** response?
 - f. How frequently do subjects give a **H&A** response?
 - g. How frequently do subjects give a **H&P** response?

Analysis	Target			Control		
	1. <i>Some</i>	2. <i>Sometimes</i>	3. <i>Warm</i>	4. <i>Some</i>	5. <i>Sometimes</i>	6. <i>Warm</i>
3.11						
a. H	75.0%	41.7%	68.3%	84.4%	61.5%	51.0%
b. L	8.3%	5.0%	33.3%	4.2%	0.0%	43.8%
c. A	14.2%	9.2%	14.2%	74.0%	47.9%	5.2%
d. P	58.3%	28.3%	34.2%	8.3%	5.2%	2.1%
e. CS	24.2%	25.0%	31.7%	25.0%	26.0%	20.8%
f. H&A	12.5%	8.3%	11.7%	72.9%	46.9%	5.2%
g. H&P	55.5%	21.7%	34.2%	7.3%	3.1%	2.1%

Analysis 3.12

Questions:

1. Conditional on qtype=target:
 - a. Is the difference in the frequency of **H** responses in the *some* vs. *sometimes* prompts statistically significant?
 - b. Is the difference in the frequency of **H** responses in the *some* vs. *warm* prompts statistically significant?
 - c. Is the difference in the frequency of **H** responses in the *sometimes* vs. *warm* prompts statistically significant?
 - d. Is the difference in the frequency of **L** responses in the *some* vs. *sometimes* prompts statistically significant?
 - e. Is the difference in the frequency of **L** responses in the *some* vs. *warm* prompts statistically significant?
 - f. Is the difference in the frequency of **L** responses in the *sometimes* vs. *warm* prompts statistically significant?
2. Conditional on qtype=control:
 - a. Is the difference in the frequency of **H** responses in the *some* vs. *sometimes* prompts statistically significant?
 - b. Is the difference in the frequency of **H** responses in the *some* vs. *warm* prompts statistically significant?
 - c. Is the difference in the frequency of **H** responses in the *sometimes* vs. *warm* prompts statistically significant?
 - d. Is the difference in the frequency of **L** responses in the *some* vs. *sometimes* prompts statistically significant?
 - e. Is the difference in the frequency of **L** responses in the *some* vs. *warm* prompts statistically significant?
 - f. Is the difference in the frequency of **L** responses in the *sometimes* vs. *warm* prompts statistically significant?

Notes:

1. 2d) has a low expected count (2 cells (50%) have an expected count less than 5 (both have an expected count of 2)) so this result is a little less reliable than desired, and thus no strong conclusions should be drawn based on this result.

Appendix I: Data Analysis

Analysis 3.12		Chi-Square Value	<i>df</i>	<i>p</i>	Conclusion
1. Targets	a. H: <i>some</i> vs. <i>sometimes</i>	27.429	1	.000	Association $p < .001$
	b. H: <i>some</i> vs. <i>warm</i>	1.313	1	.252	Independent
	c. H: <i>sometimes</i> vs. <i>warm</i>	17.239	1	.000	Association $p < .001$
	d. L: <i>some</i> vs. <i>sometimes</i>	1.071	1	.301	Independent
	e. L: <i>some</i> vs. <i>warm</i>	22.737	1	.000	Association $p < .001$
	f. L: <i>sometimes</i> vs. <i>warm</i>	31.089	1	.000	Association $p < .001$
2. Controls	a. H: <i>some</i> vs. <i>sometimes</i>	12.765	1	.000	Association $p < .001$
	b. H: <i>some</i> vs. <i>warm</i>	24.393	1	.000	Association $p < .001$
	c. H: <i>sometimes</i> vs. <i>warm</i>	2.116	1	.146	Independent
	d. L: <i>some</i> vs. <i>sometimes</i>	4.085	1	.043	Association $p < .05$
	e. L: <i>some</i> vs. <i>warm</i>	41.282	1	.000	Association $p < .001$
	f. L: <i>sometimes</i> vs. <i>warm</i>	53.760	1	.000	Association $p < .001$

Analysis 3.13

Questions:

1. Conditional on **L**=yes, what is the frequency of Yes vs. No responses?

Analysis 3.13	Yes	No	I don't know
L = yes	19.6%	72.5%	7.8%

Analysis 3.14

Conditional on `qnum=28&response=yes`, code for a new variable:

- **F** = subject mentions that the sweater is Cindy's favorite.

Appendix I: Data Analysis

ID	F
105	0
110	0
112	0
117	1
118	1
119	0
120	1
121	0
122	1
126	1

Question:

1. Conditional on $qnum=28$ & $response=yes$, what percentage of the subjects gave an **F** response?

Analysis 3.14	F
1.	50%

Analysis 3.15

Questions:

1. Conditional on $qtype=target$:
 - a. Is the difference in the frequency of **H&P** responses in the prompts containing the scalar term *some* vs. the prompts containing *sometimes* statistically significant?
 - b. Is the difference in the frequency of **H&P** responses in the prompts containing the scalar term *some* vs. the prompts containing *warm* statistically significant?
 - a. Is the difference in the frequency of **H&P** responses in the prompts containing the scalar term *sometimes* vs. the prompts containing *warm* statistically significant?
2. Conditional on $qtype=control$:
 - a. Is the difference in the frequency of **H&A** responses in the prompts containing the scalar term *some* vs. the prompts containing *sometimes* statistically significant?
 - b. Is the difference in the frequency of **H&A** responses in the prompts containing the scalar term *some* vs. the prompts containing *warm* statistically significant?
 - c. Is the difference in the frequency of **H&A** responses in the prompts containing the scalar term *sometimes* vs. the prompts containing *warm* statistically significant?

Appendix I: Data Analysis

Analysis 3.15		Chi-Square Value	<i>df</i>	<i>p</i>	Conclusion
1. target: H&P x qtype	a. <i>some</i> vs. <i>sometimes</i>	28.202	1	.000	Association $p < .001$
	b. <i>some</i> vs. <i>warm</i>	10.540	1	.001	Association $p < .05$
	c. <i>sometimes</i> vs. <i>warm</i>	4.659	1	.031	Association $p < .05$
2. control: H&A x qtype	a. <i>some</i> vs. <i>sometimes</i>	13.552	1	.000	Association $p < .001$
	b. <i>some</i> vs. <i>warm</i>	92.444	1	.000	Association $p < .001$
	c. <i>sometimes</i> vs. <i>warm</i>	43.268	1	.000	Association $p < .001$

Analysis 3.16

Questions:

Conditional on qtype=target&scalar=warm:

1. What is the frequency of **H&P** responses in Q7?
2. What is the frequency of **H&P** responses in Q25?
3. What is the frequency of **H&P** responses in Q29?
4. What is the frequency of **H&P** responses in Q30?
5. What is the frequency of **H&P** responses in Q36?

Analysis 3.16	1. Q7	2. Q25	3. Q29	4. Q30	5. Q36
H&P	0.0%	29.2%	54.2%	66.7%	20.8%

Analysis 3.17

Questions:

1. Removing Q7 as an outlier, what is the frequency of **H&P** responses in the target prompts containing the scalar term *warm*?

Analysis 3.17	Warm Targets (minus Q7)
H&P	42.7%

Analysis 3.18

Questions:

1. Conditional on qtype=target, is the difference in the frequency of **H&P** responses in the prompts containing the scalar term *some* vs. the prompts containing *warm* (minus Q7) statistically significant?

Analysis 3.18	Chi-Square Value	<i>df</i>	<i>p</i>	Conclusion
1. target (<i>some</i> vs. <i>warm</i>): H&P x qtype	3.223	1	.073	Independent

Format:

All prompts for the experiment followed this format, and required subjects to pick either "Yes," "No," or "I don't know" as an answer.

1. The medical website says:

"If you sometimes have insomnia, then it is not a cause for concern."

Does the website mean:

If you sometimes but not always have insomnia, then it is not a cause for concern.

- Yes
- No
- I don't know

Prompts

Below are all of the prompts used in the experiment, sorted according to the type of prompt. The different types of prompts are listed below:

- **D:** Distractor prompts
- **C:** Control prompts
- **T:** Target prompts
 - **S:** prompts we judged to lean toward a strengthened interpretation
 - **U:** prompts we judged to lean toward an unstrengthened interpretation
 - **N:** "neutral" prompts; prompts that don't *a priori* seem to lean one way or the other
 - **TF:** "truth functionally" strengthened prompts; prompts that *require* a strengthened interpretation in order to be coherent

Distractor Prompts

Question 2:

Prof. Roberts has noticed some unusual trends in his students' grades this semester. He tells another faculty member:

"I'm worried that some of my students are cheating."

Does Prof. Roberts mean that he is worried that only some of his students are cheating?

Appendix II: Experimental Prompts

Question 3:

Prof. Smith held a party at his home for students interested in majoring in linguistics. He was quite pleased with the turn-out, and reported:

"I was surprised that some of the students from my intro class showed up."

Did Prof. Smith mean that some but not all of the students from his class showed up?

Question 6:

Paul is planning to attend a psychology conference in Boston in May, to which a large number of European and Asian researchers have been invited. He is disappointed to receive the following message from the conference organizers:

"We regret that some of the invited speakers are unable to attend because of visa problems."

Does the message mean that some but not all of the speakers are unable to attend?

Question 8:

Bill works as a youth councilor, taking groups of kids out for wilderness hikes. On one of his hikes, the trail went by a small lake. He later tells you:

"I was surprised that some of the kids wanted to swim."

Did Bill mean that he was surprised that only some of the kids wanted to swim?

Question 12:

Dr. Edwards is in charge of safeguarding data at the facility where he works. Some recent events have caused him to be concerned. He reports to the head of the facility:

"I'm worried that the security of some of the data might have been compromised."

Does Dr. Edwards mean that the security of some but not all of the data might have been compromised?

Question 13:

A catastrophic water main break has led to serious flooding in the Carnegie Art Museum in Pittsburgh. The museum spokesperson issues a statement saying:

"We regret that some of the artwork has been damaged or destroyed."

Does she mean that they regret that only some of the artwork has been damaged or destroyed?

Appendix II: Experimental Prompts

Question 17:

My friend Luke recently had the disturbing experience of having his house broken into. He is explaining how he came to realize that something was wrong, and says:

"I walked into my living room, and saw right away that some of the bureau drawers had been pulled out."

Does Luke mean that he saw that some but not all of the drawers had been pulled out?

Question 22:

Prof. Jones is teaching an upper level calculus class, and has just finished grading the midterm. He comes to class in a sour mood, and tells the students:

"I'm disappointed that some of you couldn't answer problem 3."

Does Prof. Jones mean that some but not all of the students couldn't answer problem 3?

Question 23:

Dr. Frank occasionally teaches a small undergraduate class in semantics. After her most recent semester, she tells one of her colleagues:

"I'm pleased that some of my students are planning to apply to grad school in linguistics."

Did Dr. Frank mean that some but not all of the students are planning to apply to grad school?

Question 26:

Mark supported several outsider candidates in the local elections in November. His candidates did better than might have been expected. He told the press:

"I'm pleased that some of my candidates were elected to office."

Did Mark mean that he is pleased that some but not all of his candidates were elected?

Appendix II: Experimental Prompts

Question 31:

My friend Jill organized a surprise birthday party for her boyfriend last week. She managed to keep everything a secret from him, and he appeared completely surprised when they came home from dinner and found a houseful of guests waiting for them. But later he told her:

"Just as you opened the door, I saw that some of the lights were on in the house."

Did he mean that some but not all of the lights were on?

Question 32:

Sandy planned a big party for her 30th birthday, and invited friends and family. As the replies to her invitations were coming in, she told you:

"I'm disappointed that some of my best friends can't come."

Did Sandy mean that she was disappointed that only some of her best friends couldn't come?

Control Prompts

Question 5:

Mary says to her mother:

"I sometimes like to sleep in."

Does Mary mean:

I sometimes but not always like to sleep in.

Question 9:

Kathy has crocheted an afghan for her new niece Alice. Kathy says:

"This blanket will keep Alice warm at night."

Does Kathy mean:

This blanket will keep Alice warm but not hot at night.

Appendix II: Experimental Prompts

Question 11:

Sally is talking to her husband Jack. She says:

"I sometimes like to drink tea at night."

Does Sally mean:

I sometimes but not always like to drink tea at night.

Question 15:

Larry and John are cleaning up after the party. Larry says:

"Max drank some of the beer."

Does Larry mean:

Max drank some but not all of the beer.

Question 19:

The web page on how to stay warm in the winter gives the following tip:

"Wearing fleece slippers around the house can help keep your feet warm."

Does the web page mean:

Wearing fleece slippers around the house can help keep your feet warm but not hot.

Question 24:

Jane raises chickens. She says to her friend Sue:

"The rooster sometimes wakes me up."

Does Jane mean:

The rooster sometimes but not always wakes me up.

Question 28:

Cindy is putting on her favorite blue sweater. She says to her mother:

"This sweater keeps me warm on cold days."

Does Cindy mean:

This sweater keeps me warm but not hot on cold days.

Appendix II: Experimental Prompts

Question 33:

The web page on how to stay warm in the winter gives the following tip:

"Wearing several thin layers of clothes will help you stay warm in cold weather."

Does the web page mean:

Wearing several thin layers of clothes will help you stay warm but not hot in cold weather.

Question 34:

Nancy and Rebecca are talking about their teenage daughters. Rebecca says:

"My daughter is sometimes hard to wake up in the morning."

Does Rebecca mean:

My daughter is sometimes but not always hard to wake up in the morning.

Question 35:

Jane has a small flock of chickens on her farm. One morning she said to her husband:

"Some of the chickens laid eggs this morning."

Does Jane mean:

Some but not all of chickens laid eggs this morning.

Question 37:

Linda is taking care of the cat. She says to her mother:

"The cat ate some of her food."

Does Linda mean:

The cat ate some but not all of her food.

Question 38:

Bob is telling his wife about his performance at the Harvest Festival. He says:

"I danced some of the dances."

Does Bob mean:

I danced some but not all of the dances.

Target Prompts

Strengthened

Question 1:

The medical website says:

"If you sometimes have insomnia, then it is not a cause for concern."

Does the website mean:

If you sometimes but not always have insomnia, then it is not a cause for concern.

Question 36:

Jim lives in England with his wife and two daughters. He and his wife are planning a trip to visit his sister in Pennsylvania. Jim says:

"If we want it to be warm when we visit, we should go in April."

Does Jim mean:

If we want it to be warm but not hot when we visit, we should go in April.

Unstrengthened

Question 4:

On the first day of class, the professor says:

"If students are sometimes late for class, then it makes me upset."

Does the professor mean:

If students are sometimes but not always late for class, then it makes me upset.

Question 18:

A professor and his TA are having a meeting. The professor says:

"If some of the students fail the exam, then I will give an extra credit assignment."

Does the professor mean:

If some but not all of his students fail the exam, then I will give an extra credit assignment.

Appendix II: Experimental Prompts

Question 30:

Laura says to her mother:

"If it is warm outside, then I don't need a sweater."

Does Laura mean:

If it is warm but not hot outside, then I don't need a sweater.

Question 39:

Bobby wants his mother to read him a story. His mother says:

"If you pick up some of your toys, then I will read you a story."

Does his mother mean:

If you pick up some but not all of your toys, then I will read you a story.

Neutral

Question 7:

Mary works in a restaurant that is famous for its pie. The restaurant menu says, "We can heat your pie!" Mary says to a customer:

"If you want your pie to be warm, please let me know."

Does Mary mean:

If you want your pie to be warm but not hot, please let me know.

Question 10:

Rachel and her husband own a sheep farm. Sometimes they have orphaned baby lambs, which need to be fed several times a day, especially when they are young. Rachel says to her husband:

"If we sometimes have lambs that need to be bottle fed, we will post it on our website."

Does Rachel mean:

If we sometimes but not always have lambs that need to be bottle fed, we will post it on our website.

Appendix II: Experimental Prompts

Question 16:

Jane is in charge of the chicken house. At the farm meeting she says:

"If foxes kill some of the chickens, then we will have less eggs to sell."

Does Jane mean:

If foxes kill some but not all of the chickens, then we will have less eggs to sell.

Question 20:

Jane wanted to increase the size of her chicken flock on her small farm, so she ordered 60 chicks of assorted breeds from an internet supplier. She reported the safe arrival of the chickens at the next farm meeting, and then said:

"If some of the chickens are black, then there will be an announcement."

Does Jane mean:

If some but not all of the chickens are black, then there will be an announcement.

Question 27:

Mary and her sister own a company that makes vegetarian frozen dinners and other pre-prepared foods. On their website, they say:

"If you sometimes have no time to cook, then our frozen dinners are a healthy alternative."

Do they mean:

If you sometimes but not always have no time to cook, then our frozen dinners are a healthy alternative.

Question 29:

The operating manual for a piece of machinery says:

"If the Z-plate is warm, turn the S-dial counter-clockwise."

Does the manual mean:

If the Z-plate is warm but not hot, turn the S-dial counter-clockwise.

Truth-Functionally Strengthened

Question 14:

Jane and Bill are planning on taking a yoga class on Thursday evenings, and need someone to watch their children. Jane says:

"If Sarah can babysit sometimes when we go to class, then Grandma can do it the rest of the time."

Does Jane mean:

If Sarah can babysit sometimes but not always when we go to class, then Grandma can do it the rest of the time.

Question 21:

A professor and her TA are having their weekly meeting. The professor says:

"If we give some of the students an extension, the others will be upset."

Does she mean:

If we give some but not all of the students an extension, the others will be upset.

Question 25:

Rob is warming up soup for his daughter's lunch. His daughter says:

"If the soup is warm then it is not too hot to eat."

Does his daughter mean:

If the soup is warm but not hot then it is not too hot to eat.