Processing Presupposition: Verifying Sentences with 'Only'

Christina Kim*

1 Introduction

1.1 Presuppositions and Sentence Verification

This paper is about presuppositions, and how the information they contain is used by comprehenders to understand what a sentence means in a given context. In particular, I'm interested in how presuppositions differ from assertions in how their content is processed procedurally. But let's start with the more basic question of why presuppositions are special to begin with. Why might we even expect that presuppositions would behave any differently from assertions, scalar implicatures, etc. in a behavioral task?

One answer to this is that presuppositions appear to have this special discourse status—they're characteristically backgrounded information, with respect to the main assertion of a sentence, which is foregrounded or 'under discussion'. Say presuppositions are noncontroversial, to borrow Stalnaker's language—to put it now in terms of 'speaker presuppositions', a presuppsition is material conveyed by a sentence that the speaker feels she can safely assume is already part of the common ground (or if it's not, is something that interlocutors won't take issue with—that is, it won't be a problem to accomodate that information or otherwise treat it as if it had already been in the common ground) (Stalnaker 1978, 2002; Simons 2002). The current study looks at this aspect of presuppositions from the point of view of processing; specifically it looks at the procedures people use to evaluate the truth or falsity of a sentence in some context, and asks where presupposition verification fits into such procedures[1].

With respect to what people do procedurally when they understand a

*Thanks to Daniel Büring, Philippe Schlenker, Christine Gunlogson, Martin Hackl, Jason Kandybowicz, and Colin Wilson for helpful discussions and comments, or for at least patiently listening to me talk about this material when it was in varying stages of unpreparedness.

1 There's also the projection behavior of presuppositions—that is, how the presuppositions of a complex sentence are related to the presuppositions of its subparts; this paper doesn't say anything about the projection problem, but cf. Chemla & Schlenker (2006), who start with different pairings of presupposition triggers and quantificational environments, and ask whether inferences predicted by various projection theories actually arise.

sentence with a presupposition, we can identify (at least) the following two states of affairs. First, it might be that presuppositions must be verified as a precondition for evaluating the truth/falsity of a sentence—in other words, you can't compute a truth value unless all presuppositions have been verified. This is a kind of direct implementation of a Heim-style update model, or any logical system that will output '#' or truth-value gaps in case of presupposition failure. I'll refer to this as the 'Preconditons' hypothesis. In contrast, according to an 'Assertion-first' hypothesis, the backgrounded status of presuppositions leads comprehenders to simply take it for granted that they are satisfied in the context, without bothering to verify it, instead giving priority to the main assertion.

The current study looks at verification procedures to pull apart the predictions of these two hypotheses.[2] Sentence-picture verification will give us data that bears on the issue of how presuppositions are (i) treated by comprehenders for the purpose of understanding the (truth-conditional) meaning of a sentence, and (ii) relatedly, how they differ from assertions in terms of representation in a mental discourse model. On the basis of three experiments, I'll argue for the following view of presuppositions and how they are processed: presuppositions are treated as backgrounded information and therefore simply assumed to be satisfied in the evaluation context. This means that when nothing else in the preceding discourse or the evaluation context independently makes the content of presuppositions salient, comprehenders can initially miss violated presuppositions. On the other hand, things that affect discourse salience or otherwise increase salience in the (visual) context will strongly influence how likely comprehenders are to notice cases of presupposition failure. While the results of Experiment 1 somewhat misleadingly suggest that comprehenders use presupposition violations immediately to reject a sentence (implying that presuppositions are verified first), Experiments 2 and 3 reveal that such situations only arise in case the content of the presupposition is made extra salient in the discourse context—that is, there is no requirement that comprehenders verify

---

2 Note that most if not all semantic theories of presupposition make no claim about the actual strategies people use when they are producing and understanding sentences. Even in dynamic approaches (Heim's context-change model; Kamp's DRT, Veltman, Zeevat, Beaver update models), where there seems to be a claim about procedure—that is, there's a sequence of steps executed in order for the context to finally be updated with the information in a sentence—these models stop short of claiming that such procedures are adopted by people as a comprehension strategy. To take a simple example, suppose that presuppositions are defined in the theory as definedness conditions on context updates (as in Heim); actual comprehenders might simply assume without verification that presuppositions are satisfied, simply because to do otherwise would yield infelicity and be totally unproductive/uninformative (cf. von Fintel 2006).

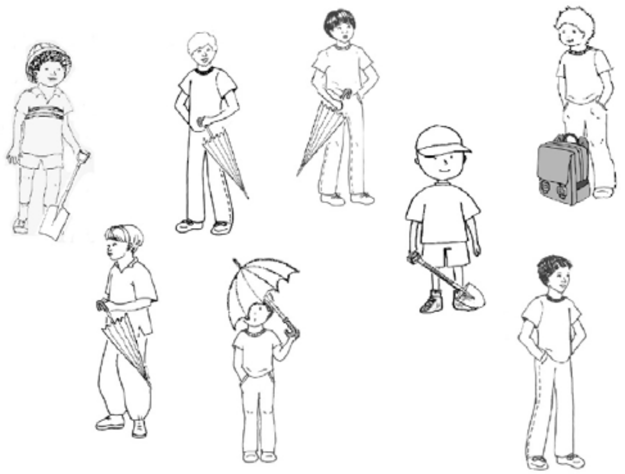

Fig. 1 Picture for *'Every kid has an umbrella'* (false).

presuppositions before evaluating the main assertion of a sentence. In fact, it will turn out that in some cases, comprehenders simply take presuppositions for granted, and only very late realize that these unverified assumptions are incompatible with information in the context.

1.2 Sentence-Picture Verification

To ask these questions we're going to have people evaluate the truth or falsity of sentences of different types, given an evaluation context (a picture). The basic idea behind this paradigm is that for each kind of sentence (really, each kind of determiner), we know what information the comprehender needs from the visual display in order to decide whether the sentence is true or false; from that, we can make some basic assumptions about how comprehenders will search a visual display for the relevant information. When visual search procedures differ with different sentence types, we assume that these different strategies will be reflected in different reaction times (RTs). We can then use subjects' RT data to infer what verification procedure they are using for some sentence type.

For instance, consider what it would take to verify or falsify the sentence in (1), given the picture in Fig. 1 as evaluation context.

(1) Every kid has an umbrella

You'd have to go through the set of individuals in the picture, checking for each one whether (s)he has an umbrella. Because the search is sequential/serial in nature, we expect sentences of this type (with universal quantification) to have relatively slow response times. Further, we expect a True/False asymmetry: False cases will be faster than True cases, since you can falsify the sentence and terminate the trial upon finding just one falsifier, whilea picture that makes the sentence true will contain no falsifiers, and therefore require exhausting the entire search space without finding one.

Now consider (2). The corresponding picture would be much like Fig. 1, except that one of the boys would have the name 'Dave' printed on his t-shirt.

(2) Dave has an umbrella

The procedure for verifying or falsifying this sentence is very different from the 'Every' case: all you have to do is find the person named in the sentence, and check whether or not he has an umbrella. Responding to such a sentence will take just as long as it takes to find the relevant individual in the picture—that is, we expect fast RTs relative to the 'Every' conditions. In addition, no True/False asymmetry of the type described above is expected.

Against the two verification strategies just described, now consider sentence (3). For now, assume that (3) asserts (4a), and presupposes (4b).

(3) Only Dave has an umbrella

(4) a. No one other than Dave has an umbrella

b. Dave has an umbrella

The two hypotheses outlined above for how presuppositions are evaluated with respect to assertions make different predictions for 'Only' sentences. On the one hand, if presuppositions are like preconditions for computing sentence meanings, then presupposition verification will be the first thing to happen, and conditions where the presupposition of 'only' is not satisfied (Dave doesn't have an umbrella) will be rejected very quickly, much like the Name conditions. But if comprehenders take presuppositions for granted and don't bother verifying them, they'll start by evaluating the assertion (whether anyone other than Dave has an umbrella), and may only discover afterward that their assumptions are at odds with the visual context (Dave doesn't have an umbrella). In this case responses would be expected to be slow, like the Every conditions (since the assertive component of an 'only' sentence is also a universal statement).

2 Experiment 1: 'Only' and Proper Names

Experiment 1 tests the three sentence types mentioned above (Every/Name/Only) in a sentence-picture verification task. Under the assumption that a universal statement (e.g. 'Every', the assertive component of 'Only') requires serial search through the visual display, and should therefore take longer than statements about individuals (e.g. the 'Name' condition, the presupposition component of 'Only'), reactions times to the Only-Presupposition failure conditions should indicate whether presupposition verification happens before evaluation of the assertion (fast RTs), or after (slow RTs).

2.1 Methods

2.1.1 Materials/Design

Sentence type (Every/Name/Only) was crossed with Truth value (True/False), with an additional Presupposition failure condition for 'Only', making a total of seven experimental conditions. Each item consisted of a sentence-picture pair; example sentences and picture descriptions are given in Table 1.

<table>
<tr><th></th><th></th><th>TRUE</th><th>FALSE</th></tr>
<tr><td rowspan="4">Sentence type</td><td>Every
Every boy has a book</td><td>8 boys, each with a book</td><td>4 boys with books; 4 with non-books</td></tr>
<tr><td>Proper name
Mark has a book</td><td>Mark with book; 7 boys with either books or non-books</td><td>Mark with a non-book; 7 boys with either books or non-books</td></tr>
<tr><td rowspan="2">Only
Only Mark has a book</td><td>7 boys wth non-books; Mark with book</td><td>4 boys with books; 3 with non-books; Mark with book</td></tr>
<tr><td colspan="2">Presupposition failure:
7 boys with non-books; Mark with non-book</td></tr>
</table>

Table 1 Example sentence-picture pairs for Experiment 1.

Picture scenes always displayed eight individuals evenly spaced out on the screen, and were created by PhotoShopping together images from verbal competence tests (Curtiss & Yamada 2004; Dunn & Dunn 1997).

Three fixed order lists were created (items were not presented randomly in order to avoid accidentally presenting adjacent items with overlapping lexical items and/or names). Each subject saw 8 items from each condition, making a total of 56 experimental items per session.

2.1.2 Procedure

26 native speakers of North American English (UCLA undergraduates) participated in the experiment.

Stimuli were presented on a Macintosh computer running PsyScope (Cohen, et al 1993). Each trial proceeded as follows: first, a sentence appeared in the center of the computer screen. After reading the sentence, subjects pressed a button, which would bring up a picture. The subject's task was to respond (by button press) 'Yes' if the sentence accurately described the picture, and 'No' if it did not. The dependent measure was the RT to the picture (i.e. the duration from the first button press to the second button press). Each session took approximately 15-20 minutes.

2.2 Results

There were main effects of both Sentence type ($F1(2,19)=44.0$, $p<.0001$; $F2(2,39)=43.3$, $p<.0001$) and Truth value ($F1(1,19)=55.3$, $p<.0001$; $F2(1,39)=54.5$, $p<.0001$), and a Sentence type-Truth value interaction ($F1(2,19)=20.3$, $p<.0001$; $F2(2,39)=20.2$, $p<.0001$)); Figure 2 shows mean

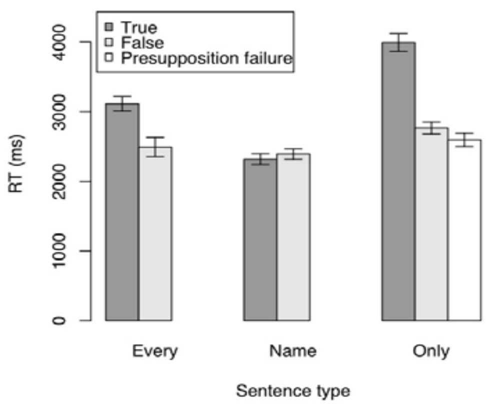

Fig. 2 Experiment 1: Mean reaction times by condition (error bars are Standard Error).

reaction times by condition.

As expected, True conditions took longer than False conditions for

'Every' ($t=3.6$, $p<.005$)[3], but not for 'Name' sentences ($t=.5$, $p<1.0$), suggesting that the RT measure is sensitive enough to pick up differences in verification procedures. Looking at the three 'Only' conditions, the Only-True condition took longer than both Only-False ($t=8.0$, $p<.0001$) and Only-Presupposition failure ($t=8.8$, $p<.0001$); Only-False did not differ from the Only-Presupposition failure condition ($t=1.3$, $p<.6$), or from the Name-True ($t=2.3$, $p<.2$) and Name-False ($t=1.7$, $p<.4$) conditions. In addition, subjects took much longer to respond to Only-True sentences than to Every-True sentences ($t=5.4$, $p<.0001$).

2.3 Discussion

The fact that the Presupposition failure condition is fast, like the Name conditions and unlike the Only-True condition, looks like support for the 'Preconditions' hypothesis: people are fast to reject cases of presupposition failure because presupposition verification precedes verifying the assertion (and since the presupposition is not satisfied in the picture, it is immediately rejected as being incompatible with the sentence).

There are reasons to be skeptical of this interpretation, however. For one thing, there's a very large RT difference between the Every-True and Only-True conditions—note that these are both universal statements that turn out true. There are only two obvious potential sources of this difference: (1) the assertive component of 'only' contains an extra negation ('for all x, it's not the case that x has some property') that isn't present in the case of 'every', and (2) 'only' possibly has this extra step of presupposition verification. But if either or both of these differences is the source of the Every-True vs. Only-True difference, the following additional data are mysterious: first, if the extra negation in the 'only' sentences is adding to the response times, Only-False should also take longer than Only-Presupposition failure, which it doesn't. Second, if the extra presupposition verification step were the source of the RT difference, Only-False should take longer than Every-False, which it doesn't ($t=1.7$, $p<.5$). We'll leave this as a puzzle for now, but return to it in Experiment 3.

There's a more basic problem with Experiment 1, which is that the 'Only' conditions involved proper names: after seeing a sentence like '*Only Dave has an umbrella*', suppose the reason subjects attended immediately to the individual labed 'Dave' in the picture is that that individual has just been named, likely bumping up the salience of both the name and the designated individual (note also that the visual presence of the label in the picture probably also contributes to its visual salience). If this is what is going on, the fact that presupposition failure is noticed immediately is simply an

---

3 P-values are adjusted throughout for multiple comparisons, where appropriate.

artifact of stimulus design—the proper name just happens to bring into attentional focus exactly that part of the picture that tells you that the presupposition is violated. To fix this confound, Experiment 2 uses definite descriptions instead of names, with the intention of making it harder to immediately identify the relevant set of individuals in the display.

3 Experiment 2: Only and Definite Descriptions

3.1 Methods

There were 8 experimental conditions, created by crossing Sentence type (Every/Def/Only#/OnlyS)[4] and Truth value (True/False); examples of sentences-picture pairs are given in Table 2. The definite description condition replaces the proper name condition from Experiment 1, and the 'only' sentences are split up into two conditions (one where the presupposition of 'only' is not satisfied, and one where it is); Experiment 2 is otherwise identical in format to Experiment 1. Pictures were adapted from the set used in Experiment 1.

Subjects saw stimuli from one of four lists. There was a total of 64 experimental trials per session. 40 UCLA undergraduates participated in the experiment.

		TRUE	**FALSE**
Sentence type	***Every*** *Every kid has a book*	8 boys/girls, each with a book	8 boys/girls: 4 with books; 4 with non-books
	Definite description *The girls have books*	2 girls with books; 6 boys with either books or non-	2 girls with non-books; 6 boys with either books or non-books
	Only—presupp not met *Only the girls have books*	2 girls with non-books; 6 boys with non-books	2 girls with non-books; 6 boys with either books or non-books
	Only—presupp satisfied *Only the girls have books*	2 girls with books; 6 boys with non-books	2 girls with books; 6 boys with either books or non-books

Table 2 Example sentence-picture pairs for Experiment 2.

4 Abbreviations for condition names: Def = Definite descriptions; Only# = Only/presupposition failure; OnlyS = Only/presupposition satisfied

3.2 Results

Mean reaction times are shown in Figure 3. As in Experiment 1, there were

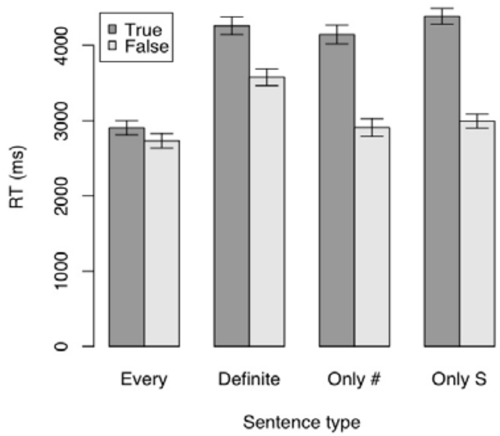

Fig. 3 Experiment 2: Mean RT by condition.

main effects of Sentence type ($F1(1,32)=130.5$, $p<.0001$; $F2(1,37)=130.9$, $p<.0001$) and Truth value ($F1(3,32)=38.7$, $p<.0001$; $F2(3,37)=38.8$, $p<.0001$), and an interaction of Sentence type and Truth value ($F1(3,38)=13.3$, $p<.0001$; $F2(3,37)=13.2$, $p<.0001$).

Post-hoc comparisons show that True conditions took longer than False conditions for both sets of Only conditions; differences were only numerical for Every and Definites. Within the Only conditions, OnlyS replicated the Experiment 1 results: True sentences took longer than False sentences ($t=10.1$, $p<.0001$). Only# show something different—the True sentences are slow, relative to both OnlyS-False ($t=7.5$, $p<.0001$) and Only#-False ($t=7.2$, $p<.0001$). OnlyS-True and Only#-True did not differ from each other ($t=1.5$, $p=.4$).

3.3 Discussion

Recall that the two verification steps that are possibly involved in verifying 'Only' sentences are (1) verifying the assertion—a negative universal statement (e.g. No one other than the girls has an umbrella), and (2) verifying the presupposition—basically identical to the Definite description case (e.g. The girls have umbrellas). In this experiment, unlike in Experiment 1, the key condition (Only#-True, where the presupposition is violated but the assertion is true) was slow relative to both Only-False conditions. This suggests that the Experiment 1 result did indeed have to do with stimulus-specific factors, as suggested, and supports the 'Assertion-first' hypothesis about processing presuppositions—there is no requirement that presuppositions be verified initially, and in the absence of external factors that bring the content of the presupposition into attentional focus, the main assertion is evaluated first.

There are a couple ways to see this in the data. First, note that Def-False < Only#-True ($t=3.4$, $p<.005$); since verifying the presupposition of 'only' will amount to the task required for the Def-True condition, the difference in RT suggests this is not the first thing that happens in the presupposition failure condition. Secondly, both OnlyS-False ($t=2.0$, $p<.2$) and Only#-False ($t=1.2$, $p=.4$) take the same amount of time as Every-False—assuming these are all universal statements that evaluate to False, it makes sense that these conditions should look similar, if in fact the universal component (i.e. the assertion) is being verified first. Finally, Def-True > OnlyS-False ($t=8.5$, $p<.0001$); if the first step were verifying the presupposition, you'd expect the reverse, since evaluating a true definite statement would be a substep of the procedure implicated for an Only sentence where the presupposition is satisfied.

Going back for a moment to Experiment 1, remember that there was a big asymmetry between the Every-True and Only-True conditions; we see the same large RT difference between the analogous two conditions in Experiment 2 (OnlyS-True > Every-True), and in fact there are similarly big differences between Every-True and the other two True conditions as well (Def-True, Only#-True). The difference between the definites and 'every' is particularly striking, since both types of sentences essentially say of some plurality of individuals that each member has some property.

Suppose the Def-True > Every-True difference actually has something to do with properties of the definite determiner (after all, definites carry presuppositions of uniqueness/maximality of their own; even if these were always satisfied in the current stimuli, they could be contributing to overall processing complexity). In this case we'd still be left with no good explanation for the other True cases. However there is another dimension on which the Every conditions differ from the rest: while domain restriction is completely trivial for 'Every' (e.g. *Every kid has a book*, where every individual in the picture is a kid), the other sentence types require that you distinguish the set of boys from the set of girls in order to be able to do the task accurately (e.g. *Only the girls have books*, where the display has both boys and girls). This means that in all but the 'Every' conditions, subjects have to apply a verification algorithm that is non-uniform, and consequently more complex[5]. Unlike the definite determiner explanation, this would give us an explanation for why 'Every' is faster than all the other conditions. The goal of Experiment 3 is to find evidence for one of these two hypotheses.

4 Experiment 3: Domain Restriction, Expectations, and Verification Complexity

In addition to the conditions from Experiment 2, Experiment 3 had an 'Every boy/girl' condition—sentences like 'Every boy has a book'. If the difference between the definites and 'every' in Experiment 2 was due to the definite determiner, 'Every boy' should pattern with 'Every kid'. On the other hand, if 'Every' was fast relative to the other conditions because only the latter required non-trivial domain restriction, 'Every boy' should have slow RTs like the Definite and 'Only' conditions, and unlike 'Every kid'.

4.1 Methods

Examples of the 5 sentence types used in Experiment 3 are in Table 3. Pictures were identical to those used in Experiment 2 (notice that the same pictures can

Sentence type	
Every kid	*Every kid has a book*
Every boy	*Every boy has a book*
Definite	*The boys have books*
Only#	*Only the boys have books*
OnlyS	*Only the boys have books*

Table 3 Sentence types and examples for Experiment 3.

be used for the 'Every boy' and 'The boys' conditions).

Subjects saw stimuli from one of five lists. With 8 tokens per cell of the experiment, there was a total of 80 experimental trials per session. 32 UCLA undergraduates participated in the experiment.

4.2 Results

5 Conceivably: given a sentence like '*Only the girls have books*', apply an algorithm like 'If boy(x), then ¬book(x)'.

Mean reaction times for Experiment 3 are shown in Figure 4. As in Experiment 2, there were main effects of Sentence type ($F1(4,22)=34.5$, $p<.0001$; $F2(4,35)=34.5<.0001$) and Truth value ($F1(1,22)=290.0$, $p<.0001$; $F2(1,35)=290.4$, $p<.0001$), and a Sentence type-Truth value interaction ($F1(4,22)=8.2$, $p<.0001$; $F2(4,35)=8.6$, $p<.0001$).

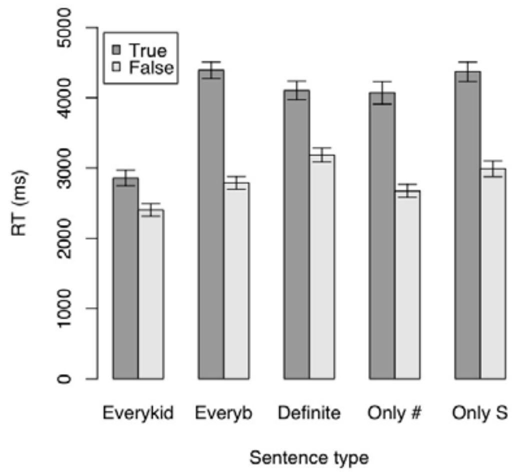

Fig. 4 Experiment 3: Mean RT by condition

True conditions took longer than False conditions across Sentence types (all adjusted $p<.001$); as in Experiment 2, the Only#-True condition was slow relative to both Only#-False ($t=8.0$, $p<.0001$) and OnlyS-False ($t=6.1$, $p<.0001$), and did not differ from OnlyS-True ($t=1.7$, $p<.1$). For the 'Every boy' conditions, True sentences pattern with Def-True sentences ($t=1.2$, $p<.25$), and took longer than 'Every kid' ($t=10.2$, $p<.0001$).

4.3 Discussion

Experiment 3 replicated the results of Experiment 2, supporting the idea that presuppositions are not explicitly verified as a first step to interpreting a sentence given some context.

The fact that 'Every boy' behaved like 'The boys' and unlike 'Every kid' suggests that the difference observed in Experiment 2 between Definites and 'Every' was not due to properties of the definite determiner, but a difference in the complexity of the verification algorithm required. The current results show that the conditions that take a long time to respond to are the ones that require subjects to distinguish among sets of individuals in the display (boys vs. girls). In contrast, the 'Every kid' cases allowed subjects to take a shortcut: even though there is domain restriction in sentences like '*Every kid has a book*', it is completely trivial since every individual in the display was a kid. Notice that it's not as simple as the fact that 'boys' and 'girls' are both necessarily 'kids'—if the types of individuals that could appear in the trials happened to include, for instance, dogs and cats, we would expect the 'Every kid'/'Every boy' asymmetry to disappear. The 'shortcut' effect we're seeing, then, has to do with subjects' expectations about what they will see from trial to trial: thus, even though the pictures provided in the experiment are the evaluation contexts for sentences in individual trials, one might think of there being a kind of 'meta'-context which reflects the set of expectations a subject develops over the set of experimental trials about what is/isn't possible in any given trial.

5 Conclusions

The three sentence-picture verification experiments presented here demonstrate, first, that information conveyed by presuppositions and assertions are used differently in sentence comprehension—they have different behavioral reflexes, as seen in Experiments 2 and 3. Further, the data support a view of processing sentence meaning in which there is no separate process that verifies that presuppositions are satisfied in the context. By virtue of being backgrounded or 'noncontroversial' material, presuppositions are taken for granted unless some contextually salient piece of information indicates otherwise. A consequence is that 'presupposition failure' arises only as a byproduct of verifying the assertive content of a sentence (as in Experiments 2 and 3), or if the presupposed material is otherwise made extra salient (e.g. visually, or due to recency of mention in the discourse, as in Experiment 1). In addition, Experiment 3 illlustrates that the complexity of sentence verification procedures is modulated by a comprehender's expectations, based on previous trials, of what should and shouldn't be a possible stimulus item (e.g. in the current experiments: pictures with girls and boys holding different types of items are possible; pictures with cats and birds, or pictures involving color, are not possible). More generally, the results reported here suggest that the actual procedure involved in verifying or falsifying a sentence in some context is never determined solely by the sentence's lexical content or syntactic structure—sentence-external things like factors influencing discourse saliency, and (possibly changing) expectations based on experience with respect to some context, affect verification procedures by changing aspects of mental discourse representations.

References

Chemla, Emmanuel & Schlenker, Philippe. 2006. Poster, OSU pragmatics workshop.
Cohen, Jonathan, MacWhinney, Brian, Flatt, Matthew & Provost, Jefferson. 1993. PsyScope: A new graphic interactive environment for designing psychology experiments. Behavioral Research Methods, Instruments, and Computers, 25(2), 257-271.
Curtiss, Susan &Yamada, Jeni. 2004. The Curtiss-Yamada Comprehensive Language Evaluation Expressive Measures (CYCLE-E). Baltimore: A Fine Line.
Dunn, Lloyd M. & Dunn, Leota M. 1997. The Peabody Picture Vocabulary Test. Circle Pines, MN: American Guidance Service.
von Fintel, Kai. 2006. What is presupposition accomodation, again?
Karttunen, Lauri. & Peters, Stanley. 1979. Conventional implicature. In: C-K Oh & D. A. Dinneen (eds.), Presupposition (Syntax and semantics 11) 1-56. New York: AP.
Simons, Mandy. 2002. Presupposition and Accomodation: Understanding the Stalnakerian Picture. Ms, CMU.
Stalnaker, Robert. 1978. Assertion. In: P. Cole, ed., Pragmatics (Syntax and semantics 9), 315-332. New York: Academic Press.
Stalnaker, Robert. 2002. Common Ground. *Linguistics & Philosophy* 25, 701-721.

Department of Linguistics
University of California, Los Angeles
Los Angeles, CA 90095
chkim@humnet.ucla.edu