

## The search for truth: Appositives weigh in\*

Margaret Kroll

*University of California, Santa Cruz*

Amanda Rysling

*University of California, Santa Cruz*

**Abstract** The semantic and pragmatic contribution of appositives to their containing sentence is a subject of continuing debate. While unidimensional semantic accounts propose that appositives contribute their truth conditions to their containing sentence, multidimensional accounts predict that they do not. In three experiments, we directly compared judgments of the truth of sentences containing appositives and sentences containing conjunctions. Our findings contribute both a methodological and a theoretical point. First, we show that no conclusions about the truth-conditional contributions of appositives can be drawn from experimental work without further investigation of how participants provide truth value judgments for complex sentences. Second, we show that while appositives appear to contribute truth values to their containing sentences, participants are highly sensitive to task features when they compute the truth value of sentences with appositives and also, crucially, with conjunctions. Specifically, we show that both sentences containing appositives and those containing conjunctions can be judged true even when the appositive or one conjunct is patently false. We conclude that it is unlikely that these results reflect semantic judgments, and suggest that they reflect truth only at the speech-act level.

**Keywords:** semantics, experimental pragmatics, truth judgments, appositives

### 1 Introduction

It is debated whether appositives, as expressing canonically not-at-issue content, contribute truth conditions to their containing sentence. The debate largely centers around the nature of the meaning that appositives contribute to their containing sentence and to the discourse. Appositives contribute entailments, but also demonstrate exceptional behavior compared to main clause content. For example, appositives generally project and contribute new, but secondary, information in discourse (Potts 2005). These properties arguably explain intuitions that (1-b) is felt to be less false

---

\* We are grateful to Adrian Brasoveanu, Pranav Anand, Donka Farkas, Jesse Harris, Dan Lassiter, Kristen Syrett, Maziar Toosarvandani, Matt Wagers, and audiences at the California Meeting on Psycholinguistics 2018, UC Berkeley's SSCircle, and SALT 29 for helpful questions and comments on content and presentation.

than (1-a), and partially motivate proposals in which appositives compose semantically separately from their containing sentence (Bach 1999; Dever 2001; Potts 2005).

- (1) a. San Francisco is the capital of California and is north of San Diego.
- b. San Francisco, which is the capital of California, is north of San Diego.

However, recent work reaffirms that the story is more complicated than merely positing that appositives compose in a separate semantic dimension. Amaral, Roberts & Smith (2007) and Harris & Potts (2009), for example, show that appositives do not always project. AnderBois, Brasoveanu & Henderson (2015) provide corpus examples showing that appositives can integrate with main clause information at the level of the discourse, and Syrett & Koev (2014) present experimental evidence that appositives are evaluated as contributing their truth conditions to their containing sentence. We follow up on these influential works in three experiments, which directly compared truth evaluations of sentences containing appositive relative clauses and conjunctions.

## 1.1 Background

Appositives are one member of a broader class of constructions that generally contribute *not-at-issue* content in a discourse. Not-at-issue content is an umbrella term that encompasses a heterogeneous set of constructions, including appositives, epithets, parentheticals, and presuppositions. There are several diagnostics that characterize constructions contributing not-at-issue content; these diagnostics are best thought of as a cluster of properties that characterize secondary information in an utterance, and not as tests that correlate tightly with particular constructions.

Not-at-issuedness is generally discussed as a property of a construction situated within a discourse.<sup>1</sup> One property of not-at-issue content is that it generally does not contribute to resolving the current Question Under Discussion, or QuD, of the discourse (Ginzburg 1996; Simons, Tonhauser, Beaver & Roberts 2010; Roberts 1996; Roberts 2012). For example, the appositive in (2) expresses information that is secondary to the QuD, which is resolved, or answered, in the main clause.

- (2) Q: What are the sailors doing?  
     A: The sailors, who see the coming storm, are sailing back into the harbor.

Not-at-issue content also generally projects past operators such as negation (Karttunen 1973; Karttunen & Peters 1979; Beaver 2001; Kadmon 2001; Potts 2005;

---

<sup>1</sup> Though see Potts (2005) for a view under which not-at-issuedness is a property of particular semantic constructions.

Simons et al. 2010; Potts 2012). However, not all not-at-issue content projects in the same manner. While presuppositions are generally blocked by ‘plugs’ such as propositional attitude verbs (Karttunen 1973), appositives can (but do not always) project beyond them (see e.g. Amaral et al. 2007; Harris & Potts 2009).

Potts (2005) influentially argued that appositives are part of a subset of not-at-issue constructions called *conventional implicatures*. The classes of constructions (including appositives, epithets, and honorifics) falling under the conventional implicature label are unified by Potts at the level of their semantic composition: conventional implicatures compose semantically independently from their containing sentence. Therefore, under the conventional implicature account, appositive relative clauses compose semantically independently from main clause content, and contribute a distinct proposition at the semantic level. This semantic independence accounts for the projection properties of appositives, as well as observations that appositives are inaccessible for binding from a main clause.

The overall picture of appositives is, however, complex. While they often exhibit the properties of not-at-issue content, they are also able at times to integrate more directly into the discourse. Potts (2005) observed that conventional implicatures act similar to main-clause content for certain discourse-level behavior, such as serving as the antecedent for inter-clausal anaphoric dependencies (see also Frazier & Clifton 2005 and Nouwen 2006) and establishing discourse referents (cf. Murray 2014). Subsequent research has also shown that appositives can sometimes respond to a QuD in the discourse and can embed under propositional attitude verbs (Amaral et al. 2007; Harris & Potts 2009; Tonhauser 2012; Koev 2013; AnderBois et al. 2015; Kroll & Wagers 2017; a.o.). Such observations are captured in the account of AnderBois et al. (2015), under which appositives and main clause content share a semantic representation, but appositives differ from main clause content by directly imposing their content on the discourse context set.

## 1.2 Current Study

The current study is indebted to the pioneering experimental work on appositives in Syrett & Koev (2014). Syrett & Koev (2014)’s results add to a building consensus that appositives are considered in the evaluation of sentential truth, that is, that appositives contribute their truth conditions to their containing sentences. Syrett & Koev (2014) found that participants overwhelmingly evaluated a sentence containing a false appositive as false. Their findings motivate an account under which appositives contribute their truth conditions to their containing sentence in a manner analogous to at-issue content, e.g. conjunctions. However, under this account, appositives are illocutionarily independent from main clause content, meaning that they contribute a separate speech act from the rest of the sentence.

We do not take exception to this account here. However, we argue that the truth-evaluation results gathered by Syrett & Koev (2014), while insightful and influential, do not themselves provide the empirical evidence needed to adjudicate in favor of or against such an account. We show that no conclusions about the truth-conditional contributions of appositives can be drawn from experimental work without further investigation of how participants provide truth value judgments for complex sentences. While we replicate Syrett & Koev (2014)'s results that appositives affect the truth judgments of their containing sentences, we demonstrate that participants are highly sensitive to task features when they compute the truth value of sentences with appositives and also, crucially, with conjunctions. In particular, we show that both sentences containing appositives and sentences containing conjunctions can be judged true even when the appositive or one conjunct is patently false. Because sentences with false conjuncts are judged true, it is unlikely that these results reflect semantic judgments, and we argue that these results should not be used as evidence for or against particular semantic analyses of appositives.

This paper presents the findings of three experiments that investigated the contributions of appositive relative clauses and conjunctions to the truth of their containing sentences. We hypothesized that judgments of truth and falsity would be affected by clausal type, such that participants would be more willing to disregard false appositives than false conjunctions in evaluations of truth. However, this hypothesis was not supported by our findings. We show instead that clausal type did *not* affect truth judgments; we found no evidence that sentences containing an appositive clause and sentences containing a corresponding conjunct clause were treated differently in truth judgments.

The experiments also investigated whether domain-general reasoning processes, such as goal-oriented behavior, influenced truth evaluations in the experimental setting. Our hypothesis was that judgments of truth and falsity would be affected by the relevance of information to conversational goals, such that information relevant to resolving a QuD would be prioritized over irrelevant information in truth evaluations. This hypothesis was borne out. The relevance of clausal information did affect truth judgments, such that relevant clauses – clauses whose information was relevant to responding to a QuD – mattered more than irrelevant clauses – clauses whose information was not relevant to responding to a QuD. These results suggest that goal-oriented reasoning can cause participants to weigh different parts of a sentence unequally in evaluations of sentential truth.

We conclude by arguing for the following specific point about appositives and a more general point about the role of experimental evaluations of truth in semantic theory. Our results show that truth judgments can be modulated by domain-general or top-down processes, such as reasoning about conversational goals. Unless we want to bake this behavior into our semantic truth-conditions – that is, unless we

want to argue that the semantic truth-conditions of conjunctions can be affected by the relevance of clausal material – then truth judgments of this nature **do not** necessarily reflect semantic truth. Therefore, the experiments presented here should not be used to adjudicate between competing theories of the semantic contribution of appositive relative clauses. More broadly speaking, it therefore follows that experimental judgments of truth as they are currently gathered are not always a good guide to theoretical semantic predictions of truth.

## 2 Experiment 1

Experiment 1 replicated and expanded upon previous experimental work on truth evaluations of appositives, notably those of [Syrett & Koev \(2014\)](#). Unlike previous experimental investigations, we directly compared sentences containing appositives with sentences containing analogous conjunctions. We also diverged from previous investigations in using basic real-world knowledge of color and shape terms for truth evaluations. Shapes were provided on the screen with target sentences, and target sentences made true or false statements about those shapes based on color and shape descriptors. Participants therefore had the evidence they needed to evaluate the sentences locally.

Another novel feature of our experimental design was to include a context sentence for each item. We reasoned that the inclusion of a context question was important, because it allowed us more control over the information status of propositions in the experimental context. The context question gave participants a guide to the local QuD, or goal, of the experimental context. Without such an explicit QuD, participants may have attempted to reconstruct this information themselves. Since we would have no way of knowing whether they had done so, or what their conclusion(s) had been, we would have faced greater uncertainty in interpreting their behavior without a preceding QuD than we did with one. In Experiment 1, the context question was maximally general, and held the same across all items and fillers: ‘What can you tell me about the shapes?’ This question was chosen so that all information contained in the target sentences could be construed as relevant in the experimental context.

In order to fully explore the behavior of sentences with appositives versus those with conjunctions, we varied the truth of clausal information in target sentences so that all possible combinations of truth and falsity were tested (see Table 1). When both clauses in a sentence expressed information that was true, we hypothesized that participants would evaluate the whole sentence as true. When both clauses expressed false information, we expected participants to evaluate the whole sentence as false. The interesting cases were those in which there was a mismatch in clausal truth, that is, when one clause in a sentence expressed true information and the other clause

expressed false information. If appositives do not contribute their truth conditions to their containing sentence, or if their content is more easily disregarded, then we expect that sentences containing a false appositive would be more likely to be evaluated as true overall than sentences containing a false conjunction. If appositives contribute their truth conditions to their containing sentence in a manner similar to conjunctions, however, then we expect to see sentences with a false appositive and sentences with a false conjunct evaluated as overall false at comparable rates.

## 2.1 Method

### 2.1.1 Participants

Forty-eight workers on Mechanical Turk participated in the experiment; all self-identified as native English speakers. Completion of the survey took an average of 52 minutes. All participants were paid \$7 for their participation.


### 2.1.2 Materials

Materials consisted of 80 experimental items comprising eight conditions. The design was  $2 \times 2 \times 2$  with the factors *clause type* (Appositive or Conjunction), *clause 1 truth* (True or False), and *clause 2 truth* (True or False). Items were presented as mini-dialogues on a single screen. Each item contained two shapes, the question prompt ‘What can you tell me about the shapes?’, and the target sentence with the options ‘True’ and ‘False’. As mentioned, a general context question was chosen in order to ensure that all parts of the item sentences would be relevant to the question.

Target sentences consisted of two clauses in which the first clause was either an appositive relative clause or a conjunct clause. If the first clause was an appositive, then the second clause was a matrix clause; if the first clause was a conjunct, then the second clause was also a conjunct clause. Examples of minimally different appositive and conjunction conditions from one experimental item are given in Table 2. Each clause in the target sentences was manipulated to be true or false relative to the shapes provided with the target sentence. An example of the appositive half of an item is given in Table 1. Shapes varied across items, and item descriptions were always basic shape or color terms that were unambiguously true or false relative to the shapes displayed.

## 2.2 Procedure

The experiment was administered on Mechanic Turk and used the IbexFarm experimental software (<http://spellout.net/ibexfarm>). Critical sentences were distributed

	
Sentence	Truth Value
The circle, which is hollow, is next to the blue triangle.	TT
The circle, which is hollow, is next to the brown triangle.	TF
The circle, which is filled, is next to the blue triangle.	FT
The circle, which is filled, is next to the brown triangle.	FF

**Table 1** Appositive half of item from Experiment 1.

Clause Type	Item
Appositive	The circle, which is hollow, is next to the blue triangle.
Conjunction	The circle is hollow and is next to the blue triangle.

**Table 2** Appositive and conjunction conditions from one item in Experiment 1.

into eight Latin Square lists and were presented in pseudo-random order. The 80 experimental items were combined with 160 filler sentences. Fillers were designed to mask the purpose of the experiment from participants. As such, they were counterbalanced for clause type, shapes, positioning of shapes, and predicates. In order to encourage participants to use the full range of the confidence scale, a combination of common and uncommon shapes was used. Subjective predicates, such as *scary*, were also included. The truth of fillers were balanced so that half of the experiment's trials were true.<sup>2</sup>

Each trial consisted of a mini-dialogue containing two shapes on the screen. Below the shapes was the context question, and below the question was a declarative sentence to be judged as "True" or "False." Participants were instructed to judge the sentence as either "True" or "False" by clicking on their choice or by using number keys. Participants were given directions containing basic examples of truth or falsity that used simple shapes and descriptions, e.g. the sentence 'A star has 5 points' is True. They were instructed to read each mini-dialogue (trial) carefully, but to answer based on their first impression of truth or falsity. After participants responded to the truth judgment task, the screen advanced to showing a scale, and participants were asked to rate their confidence about their truth judgment answer on a scale from 1 ('very unsure/unconfident') to 5 ('very sure/confident'). The screen also automatically advanced to the confidence rating screen after 5500 milliseconds if no

<sup>2</sup> Truth was, of course, subjective in some fillers and depended on the participants' choice of including appositive material. Truth was balanced in the items by assuming the inclusion of appositive content.

answer was given within that time window for the True/False task. Participants were told to rate their confidence as low if they were unsure of the truth or falsity of the previous trial's sentence. Four practice items were also given before the experiment began to familiarize the participants with the task.

Participants were given two cartoon breaks 1/3 and 2/3 of the way through the experiment. They were instructed that the experiment was paused at these points, and to take a break if they wished. At the completion of the experiment, a debriefing questionnaire and a short quiz testing color and shape competency were given.

### 2.3 Analysis

Truth judgment results were analyzed using a mixed-effects logistic regression model fit by maximum likelihood. Fixed effects were *clause type* (Appositive vs. Conjunction), *clause 1 truth* (True vs. False), *clause 2 truth* (True vs. False), and all possible interactions among them. Following Barr, Levy, Scheepers & Tily (2013), the maximal random effects structure, which included random slopes and intercepts and correlations between them by Participant and Item, was used.

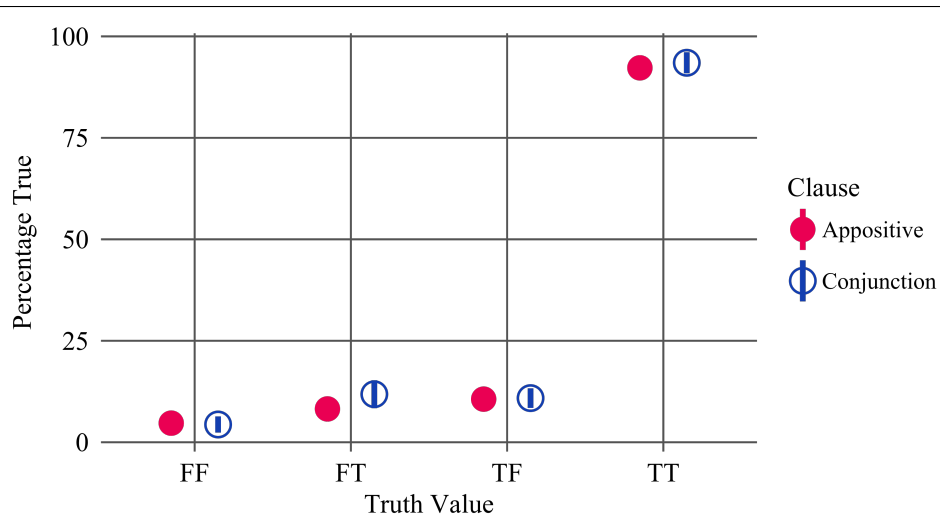
### 2.4 Results

Mean proportions of “True” judgments by clausal truth-value combinations are given in Figure 1, along with 95% confidence intervals. We predicted that sentences with a false appositive clause and a true second clause would be evaluated as “True” at a greater rate than sentences with a false conjunct and a true second clause. Counter to this prediction, we found no significant effect of *clause type* or interaction of *clause type* and *clause 1 truth*, which would have indicated that sentences with appositives were treated differently from those with conjuncts by participants. Instead, participants were sensitive only to clausal falsity, evaluating sentences with one or more false clauses as false overall. Sentences in which both clauses were true were overwhelmingly evaluated as “True.” These results were reflected in significant main effects of *clause 1 truth*, *clause 2 truth*, and an interaction between the two (all  $ps < .001$ ). No other effects reached a significance threshold of  $p < .05$ .

### 2.5 Discussion

Recall our hypothesis: if appositives do not contribute their truth conditions to their containing sentence, or if their content is more easily disregarded, then sentences containing a false appositive will be more likely to be evaluated as true overall than sentences containing a false conjunction. However, if appositives contribute their





**Figure 1** Percentage of sentences evaluated “True” by truth value combination and clause in Experiment 1. Error bars indicate 95% confidence intervals.

truth conditions to their containing sentence in a manner similar to conjunctions, then sentences with a false appositive and sentences with a false conjunct should be judged “False” at comparable rates. With this hypothesis in mind, there are two main observations to make about the results of Experiment 1. The first is that the only sentences participants evaluated as “True” were those in which both clauses expressed true information. This result demonstrates that any false clausal information in the target sentences was sufficient for the whole sentence to be evaluated as “False;” that is, participants responded in a manner consistent with the truth-table for conjunctions. The second observation is that we found no significant difference between sentences containing appositives and sentences containing conjunctions. This result shows that sentences with conjunctions and sentences with appositives were not treated differently, counter to what theories in which appositives have exceptional truth-conditional behavior predict.

One concern that could be raised regarding these results is that the use of a definite article in the target sentences results in a presupposition failure. The concern, then, is that the false information is not in fact false, but is merely a presupposition failure leading to undefinedness (Frege 1892; Strawson 1950; cf. Russell 1905). There are two main reasons why we argue that this does not change our interpretations of the results. The first reason is that theoretical predictions of the differences between falsity and presupposition failure hold that people should be *less* likely to evaluate presupposition failures as false, or they should be less confident

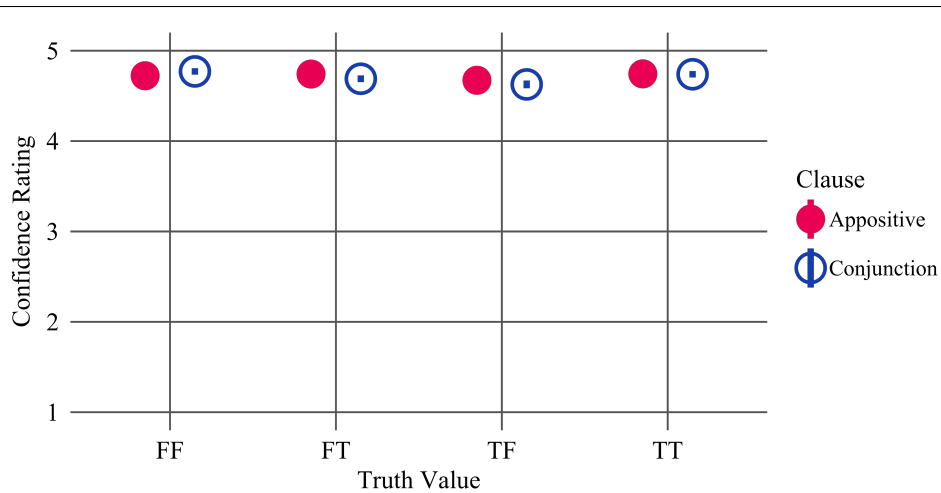
in evaluating a presupposition failure as false. We would therefore expect one, or possibly two, things to hold if this concern affected the current results: sentences with a presupposition failure should not be evaluated as “False,” or they should be evaluated as “False,” but with low confidence. However, neither of these outcomes held in our results. Sentences with one false clause were overwhelmingly evaluated as “False,” regardless of whether the false information was contained in a definite description, as shown in Figure 1.<sup>3</sup> As shown in Table 1, the false information in the False-True condition did not contain a definite description, while the false information in the True-False condition did contain a definite description. Despite the presuppositional differences between the two clauses, we found no difference in how the two clauses were evaluated.

Theoretical views under which presupposition failures would affect the current results also predict that participants would be less confident of evaluating sentences with presupposition failures as false. However, Figure 2 shows that participants were in fact very highly confident of their evaluations in each condition. We therefore found no evidence that our results are confounded by the use of a definite description in the second clause. Instead, our results fit neatly into a growing body of theoretical and experimental work on the presuppositions of definite descriptions that show that multiple factors, such as topicality and verifiability, affect intuitions of truth and falsity. We refer the interested reader to these works (Strawson 1964; Lasersohn 1993; von Stechow 2004; Schoubye 2010; Abrusán & Szendrői 2013; Schwarz 2014; Schwarz 2015).

Another question that could be raised at this point is whether the decision to have a forced-choice task, in which participants must choose either True or False for each sentence, could have affected our findings and therefore our conclusions about the experiment. For example, we could have chosen to give participants a scale of truth and asked them to rate sentences in that manner. While both options are valid methodologies, we do not believe our choice of methodology has influenced our conclusions in any interesting way. Previous studies have shown that providing a forced-choice task with a confidence rating yields the same results as presenting a single scalar task (Macmillan & Creelman 1991). Furthermore, Syrett & Koev (2014) used a scalar truth task in their Experiment 5, and found that participants judged sentences with false appositives at the False-most end of the scale. Because previous studies have shown equivalency between the two types of task, and because we replicated Syrett & Koev (2014)’s results in which a scale of Truth/Falsity was used, we argue that our particular choice of methodology has not affected our conclusions in any interesting or notable way.

---

<sup>3</sup> Note that while the means are not at 0% True, they are at what is reasonably considered to be floor. Completely categorical results are rarely if ever collected in experimental settings, due to individual variation and trial errors.



**Figure 2** Mean confidence ratings for Experiment 1 by truth value combination and clause. Error bars indicate standard error of the mean.


### 3 Experiment 2

In Section 1 we noted that appositives, as canonically contributing not-at-issue content, generally do not contain information relevant to resolving the current QuD. In Experiment 1, we used a general QuD such that both clauses in the experimental items could be construed as containing relevant information. However, this type of general QuD may have encouraged participants to consider appositive content in their evaluation of the sentences' truth or falsity. In Experiment 2, we used a specific QuD that targeted only information in the second clause of the experimental items. Therefore, only the matrix clauses/2nd conjuncts had information that could be construed as relevant to answering the QuD. Our hypothesis was that a specific QuD would encourage participants to disregard the irrelevant appositive content in their evaluations of sentential truth and falsity.

#### 3.1 Method

##### 3.1.1 Participants

Forty-eight workers on Mechanical Turk participated in the experiment; all self-identified as native English speakers. Completion of the survey took an average of 50 minutes. All participants were paid \$7 for their participation.

	
<b>What color is the circle?</b>	
<b>Sentence</b>	<b>Truth Value</b>
The circle, which is next to the blue triangle, is green.	TT
The circle, which is next to the blue triangle, is brown.	TF
The circle, which is next to the pink triangle, is green.	FT
The circle, which is next to the pink triangle, is brown.	FF

**Table 3** Appositive partial item from Experiment 2 with specific question targeting matrix clause content.

### 3.1.2 Materials

The materials for Experiment 2 were the same as the materials for Experiment 1; however, instead of the single question prompt ‘What can you tell me about the shapes?’, each item had a different question. The questions in Experiment 2 targeted *only* the information given in the second clause of the target sentences, which is the matrix clause in the appositive conditions and the second conjunct in the conjunction conditions. A partial item is given in Table 3. Questions in the fillers included a balanced mix of general and specific questions that targeted different information in the filler sentences. This variation was included so that participants could not learn which information the question was going to target, or, conversely, where they could find the information relevant to the QuD in the sentences.

### 3.2 Procedure

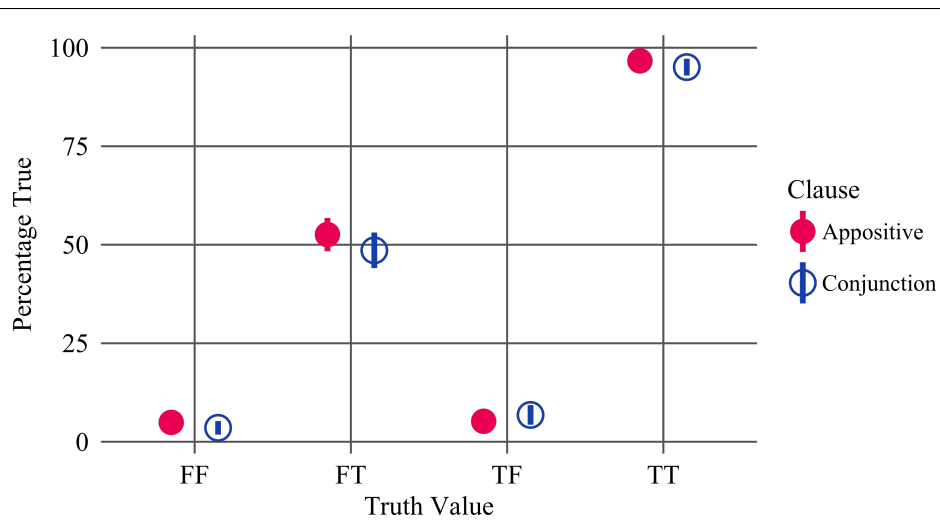
The procedure for Experiment 2 was identical to that for Experiment 1.

### 3.3 Analysis

Truth judgment results were analyzed in the same way as those for Experiment 1, except that the random effects structure included only uncorrelated random slopes and intercepts by Participant and Item.

### 3.4 Results

Mean proportions of “True” judgments by clausal truth-value combinations are given in Figure 3, along with 95% confidence intervals. For Experiment 2, we predicted



**Figure 3** Percentage of sentences evaluated “True” by truth value combination and clause in Experiment 2. Error bars indicate 95% confidence intervals.

that the specific QuD in the experimental context would encourage participants to disregard false appositives at a greater rate than in Experiment 1, leading to more sentences being evaluated as true when the sentence contained a false appositive but a true main clause. This hypothesis is borne out in the results. As shown in Figure 3, the False-True appositive condition sentences were judged “True” at a mean rate of 53%, up from 8% for Experiment 1. However, the False-True conjunction condition sentences were similarly judged “True” at a mean rate of 49%, up from 2% for Experiment 1. This finding for the conjunction condition was not predicted. The increase in the number of sentences evaluated as “True” in the False-True condition in both the appositive and conjunction conditions is reflected in our failure to find any statistically significant differences between the two clause types. In each truth value combination, sentences with an appositive first clause and sentences with a conjunct first clause patterned together.

These results were reflected in a significant main effect of *clause 2 truth* and an interaction between the *clause 1 truth* and *clause 2 truth* (both  $ps < .001$ ). No other effects reached a significance threshold of  $p < .05$ .

### 3.5 Discussion

We hypothesized that a specific QuD targeting only the second clause in the target items would encourage participants to disregard the irrelevant appositive information.

There are two main things to notice in the results in Figure 3. As we found in Experiment 1, sentences with two true clauses were overwhelmingly evaluated as “True,” and sentences with two false clauses were again overwhelmingly evaluated as “False.” The key difference between Experiment 2 and Experiment 1 is that we now found sentences in the False-True condition evaluated as “True” about 50% of the time in Experiment 2. This means that, in sentences with a false irrelevant clause and a true relevant clause, participants were willing to disregard, or ignore, the false information about half of the time. Note that this finding cannot be due simply to a mismatch in truth value, as the True-False condition – in which the irrelevant clause contained true information and the relevant clause contained false information – was again evaluated as False overall. We see then that participants are sensitive to clausal relevance, namely that they are willing to disregard false irrelevant information, but are not willing to disregard false relevant information.

Importantly, we again found no difference between sentences containing appositives and sentences containing conjunctions. This means that participants overall disregarded false irrelevant conjunct clauses at the same rate that they disregarded false irrelevant appositive clauses. We also found again that participants were highly confident in their decisions, as shown in Figure 4. Although there is a slight dip in the confidence ratings of the False-True condition, the mean ratings of the condition are still high on the scale, and we see no significant difference between the confidence ratings of the appositive condition and the conjunction condition.<sup>4</sup>

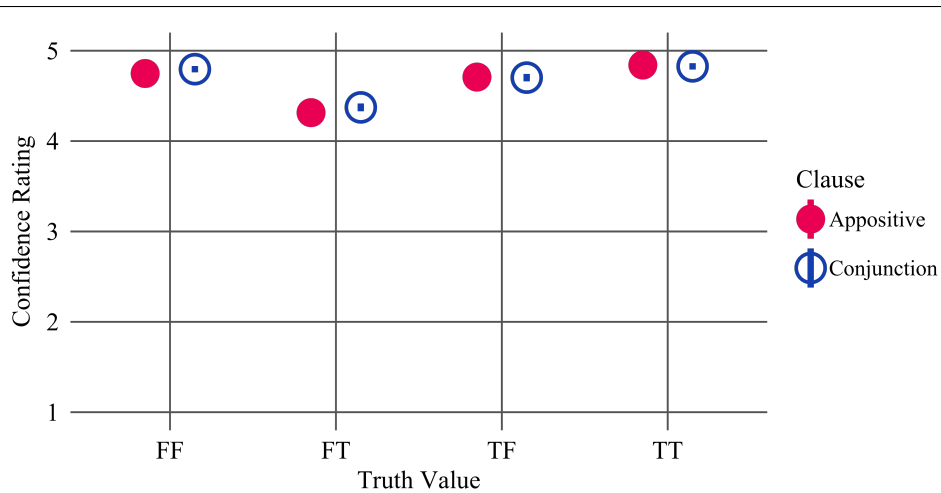
Could the results of Experiment 2 be due to participants selectively allocating attention to relevant portions of the sentences? That is, could participants have been scanning the sentences in order to find the relevant information, and therefore not fully reading or understanding the full sentences? If participants who responded “True” in the False-True condition in Experiment 2 did so because they were not fully reading the first clause, then we predict that participants responding “True” will perform worse on a verbatim memory recognition test – which requires participants to recognize material from the sentence they just read – than participants responding “False.” Experiment 3 replicated the design of Experiment 2, but included a verbatim memory recognition task that probed this question.

#### **4 Experiment 3**

Experiment 3 had the same design as Experiment 2; however, a verbatim memory recognition task was included in order to test the hypothesis that participants who responded “True” in the False-True condition in Experiment 2 were not fully reading the sentences.

---

<sup>4</sup> A maximal ordinal logistic regression model with full random effects structure showed no significant effect of clause.



**Figure 4** Mean confidence ratings for Experiment 2 by truth value combination and clause. Error bars indicate standard error of the mean.

## 4.1 Method

### 4.1.1 Participants


Forty-eight workers on Mechanical Turk participated in the experiment; all self-identified as native English speakers. Completion of the survey took an average of 55 minutes. All participants were paid \$8 for their participation.

### 4.1.2 Materials

The materials for Experiment 3 were the same as the materials for Experiment 2; however, after 1/4 of the trials participants were required to answer a comprehension question. Each participant saw 60 total comprehension questions, balanced across all item conditions and fillers. Comprehension questions were also balanced for truth, so that 50% of the questions were true and 50% were false. If the question was false, it asked about a shape that was shown on the screen but was not mentioned in the target sentence. Therefore, participants could not answer correctly based only on what shapes were shown on the screen (as opposed to what was mentioned in the target sentences).

Items with comprehension questions were randomly distributed throughout the experiment, and participants did not know which trials would have questions. On trials with comprehension questions, a screen would appear after submission of an item's confidence rating asking the participant to answer 'Yes' or 'No' to a

---


<b>What color is the circle?</b>
The circle, which is next to the brown triangle, is green.
Did the sentence mention a blue triangle?

**Table 4** Example post-rating comprehension question from Experiment 3.

---

comprehension question about the item they just completed. An example trial is given in Table 4.

## 4.2 Procedure

The procedure for Experiment 3 was identical to Experiments 1 and 2, but included additional training and example items for the comprehension question task.

## 4.3 Analysis

Truth judgment results were analyzed in the same way as those for Experiment 2.

## 4.4 Results

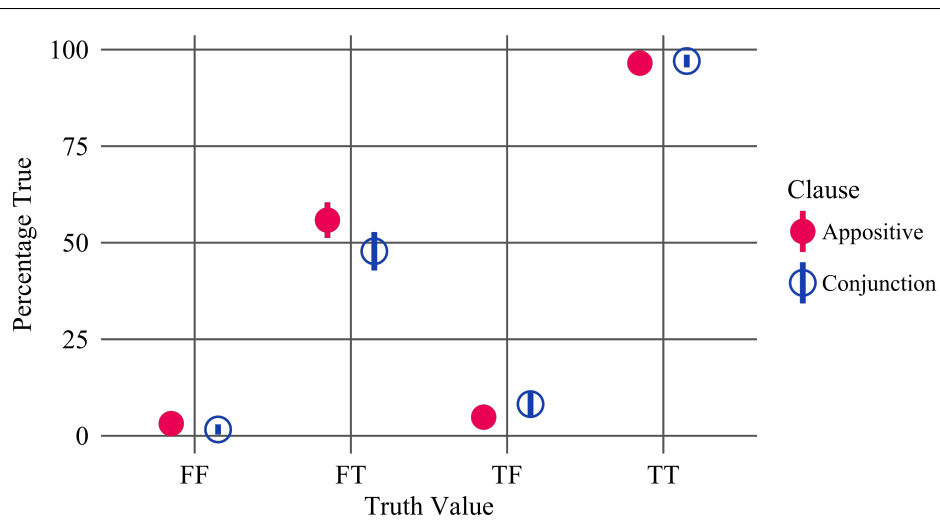
Experiment 3 replicated the results of Experiment 2. As shown in Figure 5, the False-True appositive condition sentences for both the appositive and conjunction conditions were again judged “True” at a mean rate of about 50%. We also again found no statistically significant difference between the appositive condition and the conjunction condition; in each truth value combination, sentences with an appositive first clause and sentences with a conjunct first clause patterned together.

These results were reflected in a main effect of *clause 2 truth*, and an interaction of *clause 1 truth* and *clause 2 truth* (both  $ps < .001$ ). No other effects reached a significance threshold of  $p < .05$ .

## 4.5 Discussion

We predicted that if participants who responded “True” in the False-True condition did so because they were not fully reading the first clause, then those participants would perform worse on a verbatim memory recognition task than participants who responded “False” in the False-True condition. This hypothesis is based on the proposal that, if participants were merely scanning the sentences for the relevant



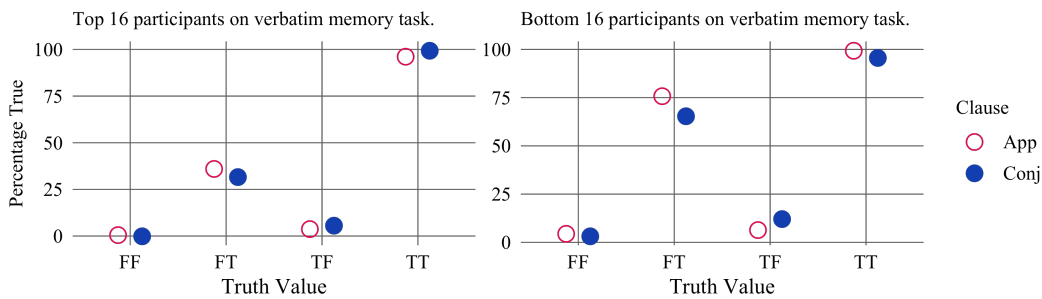


**Figure 5** Mean percent of sentences evaluated True by truth value combination and clause in Experiment 3. Error bars indicate 95% confidence intervals.

information to respond to the QuD, then their recognition of sentential material would be poorer than participants who were fully reading each sentence. Figure 6 shows the truth evaluations of participants broken down by their overall performance on the verbatim memory task. As hypothesized, the bottom 1/3 scoring participants were more likely to evaluate the False-True condition as “True” than the top 1/3 scoring participants. This may appear at first glance to provide evidence for the proposal that participants responding “True” were not fully reading the target sentences. However, overall accuracy scores suggest that this is not the case. Overall mean (and median) accuracy was 82% correct (sd=12%), and all participants scored above chance on what was a challenging task. Because participants had such high accuracy overall, it is unlikely that the findings in Experiments 2 and 3 can be explained solely by participants not fully reading the experimental sentences.

## 5 General Discussion

Our initial hypothesis was that judgments of truth and falsity would be affected by clausal type and by the relevance of propositional information to conversational goals. Contrary to expectations, we found that judgments of truth and falsity were not affected by clausal type: participants were not more willing to disregard false appositives than false conjunctions in truth evaluations. However, we found that judgments of truth and falsity were affected by the relevance of information to



**Figure 6** Mean percent of sentences evaluated True in Experiment 3 by participant score on verbatim memory task.

conversational goals: information relevant to resolving a QuD was prioritized over irrelevant information in evaluations of truth.

Our surprising finding is that even evaluations of conjunctions – whose semantic truth-conditional contributions are not under debate – could be pushed around in experimental settings. Do we want to let go of our basic beliefs about the truth conditional contributions of conjunctions? That is, do we want to propose that the semantic truth conditions of conjunctions are sensitive to clausal relevance? We argue that we do not. Therefore, we should take seriously the claim that experimental investigations of truth can tell us something about language. But, we should be cautious in how we argue from experimental results to theoretical implications.

What can we conclude, then, from our experimental findings here, and from experimental investigations of truth more generally? First, our findings reinforce the idea that there is no strict mapping from speaker evaluations of truth to semantic truth. Results of this nature do tell us when people in a particular situation are willing to apply the predicate ‘true’ to a sentence. Experimental investigations of truth such as those presented here also provide insight into how judgments of truth can be modulated by top-down, goal-driven behavior (see also [von Stechow 2004](#) and [Abrusán & Szendrői 2013](#)).

Finally, we conclude by restating what these results can and cannot tell us about appositives. Our results suggest that appositives contribute to evaluations of truth at the level of the discourse. Our results do not, however, adjudicate between theories about the semantic, truth-conditional contribution of appositives to their containing sentence. The relevance of the experimental results of truth-evaluation studies to semantic theory remains an open question.

## References

- Abrusán, Márta & Kriszta Szendrői. 2013. Experimenting with the King of France: Topics, verifiability, and definite descriptions. *Semantics and Pragmatics* 6(10). 1–43.
- Amaral, Patricia, Craige Roberts & E. Allyn Smith. 2007. Review of *The Logic of Conventional Implicatures* by Chris Potts. *Linguistics and Philosophy* 30. 707–749.
- AnderBois, Scott, Adrian Brasoveanu & Robert Henderson. 2015. At-issue proposals and appositive impositions in discourse. *Journal of Semantics* 32. 93–138.
- Bach, Kent. 1999. The myth of conventional implicature. *Linguistics and Philosophy* 22. 327–366.
- Barr, Dale J., Roger Levy, Christoph Scheepers & Harry J. Tily. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68. 255–278.
- Beaver, David. 2001. *Presupposition and Assertion in Dynamic Semantics*. CSLI Publications.
- Dever, John. 2001. Complex demonstratives. *Linguistics and Philosophy* 24(3). 271–330.
- von Fintel, Kai. 2004. Would you believe it? The King of France is back! (Presuppositions and truth-value intuitions). In Mara Reimer & Anne Bezuidenhout (eds.), *Descriptions and beyond*, 315–341. Oxford: Oxford University Press.
- Frazier, Lyn & Charles Clifton. 2005. The syntax-discourse divide: Processing ellipsis. *Syntax* 8. 121–174.
- Frege, Gottlob. 1892. Über sinn und bedeutung. In B. McGuinness (ed.), *Zeitschrift für Philosophie und Philosophische Kritik*, Oxford: Basil Blackwell.
- Ginzburg, Jonathan. 1996. Dynamics and the semantics of dialog. In J. Seligman (ed.), *Language, Logic, and Computation, Volume 1*, Stanford: CSLI.
- Harris, Jesse A. & Christopher Potts. 2009. Perspective-shifting with appositives and expressives. *Linguistics and Philosophy* 32. 523–552.
- Kadmon, Nirit. 2001. *Formal Pragmatics*. Blackwell Publishers Inc.
- Karttunen, Lauri. 1973. Presuppositions of compound sentences. *Linguistic Inquiry* 4. 169–193.
- Karttunen, Lauri & Stanley Peters. 1979. Conventional implicature. In C.K. Oh & D.A. Dinneen (eds.), *Syntax and Semantics 11: Presupposition*, Academic Press: New York.
- Koev, Todor. 2013. *Apposition and the Structure of Discourse*: Rutgers University - New Brunswick PhD dissertation.
- Kroll, Margaret & Matt Wagers. 2017. Is working memory sensitive to discourse status? Experimental evidence from responsive appositives. *XPrag* .

- Laserson, Peter. 1993. Existence presuppositions and background knowledge. *Journal of Semantics* 10(2). 113–122.
- Macmillan, Neil A. & C. Douglas Creelman. 1991. *Detection Theory: A User's Guide*. Cambridge University Press.
- Murray, Sarah E. 2014. Varieties of update. *Semantics and Pragmatics* 7(2). 1–53. doi:10.3765/sp.7.2.
- Nouwen, Rick. 2006. On appositives and dynamic binding. *Research on Language and Computation* 0. 1–18.
- Potts, Christopher. 2005. *The Logic of Conventional Implicature*. Oxford University Press.
- Potts, Christopher. 2012. Conventional implicature and expressive content. In Claudia Maienborn, Klaus von Stechow & Paul Portner (eds.), *Semantics: An international handbook of natural language meaning*, vol. 3, 2516–2536. Berlin: Mouton de Gruyter.
- Roberts, Craige. 1996. Information structure in discourse. *OSU Working Papers in Linguistics* 49. 91–136.
- Roberts, Craige. 2012. Information structure in discourse: Towards an integrated formal theory of pragmatics. *Semantics and Pragmatics* 5(6). 1–69. doi:10.3765/sp.5.6.
- Russell, Bertrand. 1905. On denoting. *Mind* 14. 479–493.
- Schoubye, Anders J. 2010. Descriptions, truth value intuitions, and questions. *Linguistics and Philosophy* 32(6). 583–617.
- Schwarz, Florian. 2014. False but slow: Rejecting statements with non-referring definites. Ms., University of Pennsylvania.
- Schwarz, Florian. 2015. *Experimental Perspectives on Presuppositions*. Springer.
- Simons, Mandy, Judith Tonhauser, David Beaver & Craige Roberts. 2010. What projects and why. In Nan Li & David Lutz (eds.), *Semantics and Linguistic Theory (SALT) 20*, CLC.
- Strawson, Peter F. 1950. On referring. *Mind* 59. 320–344.
- Strawson, Peter F. 1964. Intention and convention in speech acts. *The Philosophical Review* 73(4). 439–460.
- Syrett, Kristen & Todor Koev. 2014. Experimental evidence for the truth conditional contribution and shifting information status of appositives. *Journal of Semantics* 0. 1–53.
- Tonhauser, Judith. 2012. Diagnosing (not-) at-issue content. In Elizabeth Bogal-Allbritten (ed.), *Sixth meeting on the semantics of under-represented languages of the Americas (SULA 6) and SULA-Bar*, 239–254.

Kroll and Rysling

Margaret Kroll  
Stevenson College  
1156 High Street  
Santa Cruz, CA 95064  
[margaretkroll@ucsc.edu](mailto:margaretkroll@ucsc.edu)

Amanda Rysling  
Stevenson College  
1156 High Street  
Santa Cruz, CA 95064  
[rysling@ucsc.edu](mailto:rysling@ucsc.edu)