

YALE UNIVERSITY

Mind the (Truth-Value) Gap: truth, falsity, and everything in between¹

April 30, 2013

Author:
Paulina HADUONG

Advisers:
Professor Laurence HORN
Professor Joshua KNOBE

Presuppositions and assertions have been studied by pragmaticists and semanticists alike. When the presuppositions associated with a particular assertion are true, a truth-value judgment of the assertion is fairly straightforward. When an assertion triggers a false presupposition, however, the judgment of such an assertion becomes a little more complicated, and when that assertion is itself negated... How are we to think about such assertions? Chemla and Bott (2010), using data from a truth-verification task, argued that subjects tended to derive presuppositions in a global-first manner, but their experimental paradigm failed to take into account the possible existence of a truth-value gap. Using stimuli adapted from Chemla and Bott (2010) and two different experimental paradigms, data collected via Amazon Mechanical Turk indicates that a truth-value gap may, indeed, exist, depending on certain pragmatic factors. Results suggest that we should continue to explore more non-bivalent accounts in future experimental work.

¹*Submitted to the faculty of the Department of Linguistics in partial fulfillment of the requirements for the degree of Bachelor of Arts*

Many thanks to Larry Horn, Josh Knobe, Zoltán Gendler Szabó, Maria Pinango, Raffaella Zanuttini, and LING 490/491 (aka Awkward Gloss) for their assistance, advice, and support. Funding for this project was provided by the Berkeley College Commonplace Society (Mellon Forum) Undergraduate Research Grant.

Contents

1. Introduction	4
2. What is a (Pragmatic) Presupposition?	4
2.1. Presupposition Triggers	4
2.2. The potential for presupposition failure	5
2.2.1. Accommodation	6
2.2.2. Cooperation	7
2.2.3. Non-catastrophic presupposition failure	7
3. The Projection Problem	8
3.1. The semantic perspective	8
3.2. The pragmatic perspective	9
4. Of Gaps and Gluts	11
4.1. The king of France, once more	11
4.2. A further examination of gluts	11
4.3. Bullshit?	12
5. Experimental Work	13
5.1. Presupposition projection processing	13
5.2. Experimenting with the king of France	15
5.3. Fat Tony might be dead	17
6. Experimental Design	18
6.1. What is Amazon Mechanical Turk?	18
6.1.1. The demographics	18
6.1.2. Participant compensation	19
6.1.3. Participant recruitment and data quality	19
6.1.4. Using MTurk for experimental research	19
6.2. Stimuli	20
7. Experiment 1	21
7.1. Participants	21
7.2. Design	21
7.3. Procedure	22
7.4. Results	22
7.4.1. Data treatment	22
7.4.2. Comparing to Chemla and Bott’s Results	22
7.5. Discussion	23
8. Experiment 2	24
8.1. Participants	24
8.2. Design	25
8.3. Procedure	25
8.4. Results	26
8.4.1. Data treatment	26

8.4.2. Choice Proportions	26
8.5. Discussion	27
9. Experiment 3	28
9.1. Participants	28
9.2. Design	28
9.3. Procedure	28
9.4. Results	28
9.4.1. Data treatment	28
9.4.2. Choice Proportions	29
9.5. Discussion	30
10. Conclusion	30
10.1. Next steps	32
A. Appendix: MTurk and Qualtrics	33
A.1. MTurk Demographics: a closer look	33
A.2. Recruiting and compensating MTurk <i>workers</i>	35
A.3. Notes on Using Qualtrics with MTurk	35
A.3.1. Several surveys, one block at a time	35
A.3.2. Screening Survey	36
A.3.3. One survey, in randomized blocks	36
B. Appendix: Comments by Participants	36
C. Appendix: Study Forms	37
C.1. Informed Consent	37
C.2. Cover Story	37
C.3. Stimuli	38

1. Introduction

For many years, a unified theory of presuppositions and their properties has been sought by both semanticists and pragmaticists. Current questions in the field include the projection problem (the interaction of presuppositions with various lexical operators) and the origin problem (an exact typology of presupposition triggers).

Emmanuel Chemla and Lewis Bott, in Chemla and Bott (2010), based on certain predictions made by the dynamic semantic and pragmatic model, discuss the implications of their experiments for a cognitive processing model of presuppositions, but they subscribed to a bivalent truth-value framework. Experiments discussed in this paper suggest that perhaps we ought to take truth-value gap theories under careful consideration, before we can begin to make predictions about any cognitive processing models.

In this paper, I begin by discussing the philosophical motivations behind both the semantic and pragmatic viewpoints, before giving an overview of current experimental work. Afterwards, I give an account of the experimental paradigms I explored in various experiments, before explaining the results from those experiments, and discussing the implications for our theories about presuppositions.

2. What is a (Pragmatic) Presupposition?

In conversation, we take many things for granted, which is to say that we often *presuppose* information.² These things taken for granted could be considered to be “backgrounded,” or to be the information contained within a sentence which is not “at issue.” For instance, suppose Alice and John are at an office party. Alice says, to John:

- (1) Sarah is also drinking a martini.
 - a. There is a person named Sarah at the party.
 - b. Michael is drinking a martini at the party.

Here, Alice asserts that Sarah is drinking a martini, but she also presupposes several things. For one, Alice presupposes (1a), as well as presupposing that she and John share (1a) (or else, John would not know to whom Alice was referring). For another, both she and John presuppose (1b).

In order for assertion (1) to succeed, it appears as if John and Alice must share certain presuppositions. But what exactly are presuppositions, and from where do they originate? Do all presuppositions need to be shared in order for an assertion to succeed? This section will seek to address the following: when presuppositions arise, when they fail, and possible repair strategies.

²This is a pragmatic view of presuppositions. From a semantic perspective, sentences, and not speakers, presuppose information.

2.1. Presupposition Triggers

It is widely understood that presuppositions can be “triggered” by certain constructions, though the exact typology of these presupposition triggers is still not defined (Kiparsky and Kiparsky (1970), Simons (2007), Strawson (1950)). For instance, in (1), the word *also* triggers (1b). A few examples of lexical constructions and classes are listed below:

- (2) The President of the United States is standing in the corner. *definite description*
p: The United States has a president.
- (3) Barack Obama is standing in the corner. *names*
p: Barack Obama exists.
- (4) James knows Obama won a second election. *factives*
p: Obama won a second election.
- (5) Obama has stopped smoking. *aspectual*
p: Obama has smoked before.
- (6) All Obama’s children are girls. *quantifiers*
p: Obama has children.

In each case, the assertion’s presupposition *p* seems to be necessary for the success of the assertion. If the United States did not have a president, than who does (2) pick out? Who is standing in the corner? When we make conversation, in keeping with Gricean maxims, speakers (in the standard case) strive to make assertions that are true and relevant (among other qualities). Consequently, it seems important that the backgrounded information (the presuppositions) must also be true. But do presuppositions necessarily need to be shared? What happens when a presupposition fails, whether that failure is because it is false, or because it is missing from the conversation’s common ground?

2.2. The potential for presupposition failure

For instance, when Alice asserted (1), what if Sarah did not exist? Another example, the classic example, is

- (7) The king of France is bald.

In the standard literature, the failure of a presupposition appears to be a catastrophic event, which is to say that it halts the conversation. If (1a) fails, and Sarah (whom John knows) is at home instead, or John doesn’t know Sarah, John might ask, “Who are you talking about?” Alternatively, if John sees that Sarah is the only one drinking a martini at the party, he might say, “Who *else* is drinking a martini?” If John knew that Alice was wrong because he saw Sarah drinking a gin and tonic, he would likely reject her assertion by saying, “No, that’s not true,” or possibly “No, she’s drinking a gin and tonic.”³

These rejections are different from the questions for clarification, but regardless, “the whole assertive enterprise is wrecked by the failure of [the speaker’s] presupposition”

³For a more systematic treatment of “denial particles,” please see Smith et al. (2013).

(Strawson, 1964). However, it has been observed that although the presupposition can fail, or is missing, it's not clear that the assertion necessarily fails. For instance, Alice could have said,^s

(8) The woman who is also drinking a martini is Sarah.

Now, even if John knew that the woman Alice was pointing out was drinking a gin and tonic, that does not necessarily mean that Alice failed in her assertion. “The woman who is (also) drinking a martini” is a referential use of a definite description (as opposed to attributive) and even if the description is not wholly accurate, Alice has still succeeded in picking someone out (Donnellan, 1966).

The *false presupposition* does not disrupt the conversation.

Alternatively, Alice could have said,

(9) My sister Sarah is the one also drinking a martini.

From all appearances, when Alice asserts (9), she's presupposing that she has a sister. In the standard case, if John did not also presuppose that Alice had a sister, the prediction is that (9) would fail. And yet, speakers make assertions like this one almost on a daily basis, in situations where they do not assume that their audience has all of the appropriate presuppositions in their context set at the time of utterance (Lewis, 1979). Usually these assertions go unchallenged. The *unknown presupposition*, or *informative presupposition*, creates no controversy (Von Stechow, 2008).

From these examples, it seems clear that there are instances where a listener is willing to overlook a speaker's false presupposition. There are also instances where a listener can accept (believed to be true) presuppositions which were previously not in the common ground of the conversation, updating their own context set so that the speaker and listener share the same common ground. The process by which this occurs has been called *accommodation*, and is discussed below.

2.2.1. Accommodation

Accommodation is the process by which a listener accommodates a presupposition which is not in the common ground of the conversation, which is defined as “the set of propositions that the participants in that conversation at that time mutually assume to be taken for granted and not subject to (further) discussion” (Von Stechow, 2008). To some, accommodation is a kind of repair strategy, where, “if at time t something is said that requires presupposition P to be acceptable, and if P is not presupposed just before t , then - *ceteris paribus* and within certain limits - **presupposition P comes into existence at t** ” (Lewis, 1979).⁴

Because this view of accommodation appears to be rather “magical” and “mysterious” (how does the presupposition P simply come into existence at the time of utterance?),

⁴Bolded emphasis is mine.

to others, accommodation lends credence to a dynamic framework of presupposition processing, where listeners can quickly and quietly adjust the common ground at the time the presupposition trigger is uttered, “to a new common ground that satisfies the presupposition, in a way that is most plausibly the one the speaker intended,” so long as the presupposition is uncontroversial (Von Stechow, 2008).

2.2.2. Cooperation

While the standard view of presupposition processing is one of a common ground with accommodation, where accommodation is seen as a phenomenon which must be brought into alignment with the common ground theory, Atlas suggests that “accommodation is not a peripheral notion of presupposition: it is the central notion of presupposition” (Atlas, 2005). To that end, Mandy Simons suggests that we apply the Gricean Cooperative Principle very broadly, sketching out a view without a common ground at all, where both the speaker and interpreter desire to be as cooperative as possible, accommodating unknown information whenever possible (Simons, 2007).

2.2.3. Non-catastrophic presupposition failure

Regardless of whether you subscribe to the Common Ground + Accommodation or Cooperation theory of presuppositions, there is substantial evidence that (uncontroversial) informative presuppositions can be accommodated. Under the standard view, however, false presuppositions appear to lead to catastrophic failure, and yet it is not clear that this is always the case.

- (10) The lodger next door offered me twice that sum.
There is no lodger next door.
- (11) The author of Principia Mathematica also wrote Principia Ethica.
Neither author of PM wrote PE.
- (12) All ten solar planets are inhabited.
There are eight solar planets.
- (13) The man drinking a martini is a philosopher.
The man is drinking a gin and tonic, and he is an engineer.
- (14) Jones is burning the leaves.
Smith is raking the leaves.
- (15) My cousin is not a boy anymore.
My cousin is only eight years old.
(Yablo, 2006)

For various reasons, all of these statements should strike you as false, and yet Yablo suggests that they are all false for the same reason: what they assert is false (Yablo, 2006). This is a conclusion of interest because not all of the presuppositions associated with these sentences are definitively true or false, and yet these assertions themselves have truth-values. For instance, what does *the lodger next door* pick out? Because

catastrophic presupposition failure is defined as a situation where the assertion makes no claim, *non-catastrophic presupposition failure* can be seen as occurring whenever the assertion still makes a claim. When defined in this manner, it seems clear that non-catastrophic presupposition failure does exist.

3. The Projection Problem

- (16) The king has a son.
p: There exists a king.
- (17) The king's son is bald.
p: (16) and "There exists a king."
- (18) If the king has a son, the king's son is bald.
p: (???)
(Karttunen (1973), Karttunen (1974), Heim (1983))

As we can see, although (18) carries the presupposition of (16), and the second clause of (18) is identical to (17), (18) does not actually carry the presupposition that "the king has a son." Clearly, although it can be fairly straightforward to understand the presuppositions of simple sentences, the presuppositions of a complex sentence are not necessarily the same as the presuppositions of its atomic parts (those simple sentences). This is known as the *projection problem*, or "the problem of predicting the presuppositions of complex sentences in a compositional fashion from the presuppositions of their parts" (Heim, 1983).

But not only are we interested in how to predict presuppositions, we are also invested in this question from an epistemological perspective. Once we have identified which presuppositions a complex sentence carries, we can begin to identify how those presuppositions act under linguistic operators (such as connectors), which can aid us in determining the truth-values of sentences. This process, from both a semantic and pragmatic perspective, is discussed below.

3.1. The semantic perspective

- (7) The king of France is bald.
p: There is a king of France.

At present, France does not have a monarchy, which is to say that there is no king of France. The presupposition of (7) is an existential presupposition, that "there is a king of France." Because this presupposition fails to refer, to pick anyone out, what is the resulting truth-value of (7)? According to Bertrand Russell, "one would suppose that 'the king of France is bald' ought to be nonsense; but it is not nonsense, since it is plainly false" (Russell, 1905). In a Russellian framework, when a presupposition fails, the assertion still has a truth value, whether it be *true* or *false*.

Strawson, on the other hand, counters that (7) is not nonsense, for it could be used in a significant way if "the king of France" did pick out someone in particular, and if it did, (7)

would have a truth-value. But “when we utter the sentence without in fact mentioning anybody by the use of the phrase, ‘The king of France,’ the sentence doesn’t cease to be significant: we simply *fail* to say anything true or false because we simply fail to mention anybody by this particular use of that perfectly significant phrase” (Strawson, 1950).

More recent dynamic semantics models (such as Heim (1983)) argue that the meaning of an utterance consists of two components: the presupposition(s) and the assertion. Because these two components can be teased apart, it is possible for a linguistic operator to act differently on each component (Chemla and Bott, 2010), and because this is a dynamic model, we can understand how the context is being updated constantly, which accounts for how the context, and therefore the presuppositions, of a second clause can differ from the prior clause. Prior to Heim’s dynamic framework, it was understood that assertions updated the context after each assertion, on a *global* level, but Heim’s contribution was to include an option for *local* updates: because of the availability of local contexts midway through an utterance (and thus an update), there was an availability for local accommodation as well. However, global accommodation is preferred to non-global accommodation (Heim, 1983).⁵ For Heim’s file-change semantics, the meaning of an expression can be thought of as its Context Change Potential (CCP).

Consequently, the projection problem is solved by defining the presuppositional component of the relevant operators correctly, which is determined by their truth-conditions.⁶ For example, negation as a linguistic operator can be defined in this manner, within this framework:⁷

- (19) Negation [X] = Presupposition of X *and* not [Assertion of X]
(Chemla and Bott, 2010)

Under this definition, if a negative assertion triggers a false presupposition, then the utterance is determined to be false. This will be elaborated upon in subsequent sections.

3.2. The pragmatic perspective

- (20) Moldavia is a monarchy and the king of Moldavia is powerful.
(21) # The king of Moldavia is powerful and Moldavia is a monarchy.

In pragmatic theories, speakers, not sentences, have presuppositions. “Presupposition involves not only taking the truth of something for granted, but also assuming that others do the same” (Stalnaker, 1999). In a pragmatic theory of presuppositions and

⁵This is known as the PGA: preference for global accommodation.

⁶One of the reasons the semantic theories come under fire is due to the fact that they are capable of generalizing but lack explanatory force (Schlenker, 2008).

⁷There are several different ways to define negation, but since Chemla and Bott (2010) define it in this manner, and Experiments 1-3 are primarily based off their work, I am using this one. It’s true that this kind of conjunction doesn’t explain why the first “conjunct” isn’t at issue while the second one is, but before processing the negation, I suppose that one would have to have already defined what was at-issue (the content of the assertion) and what was not (the presupposition).

assertions, speakers assume that they share a context set, or a common ground, with their addressees. When a speaker asserts something, they are aiming “to reduce the context set in a certain determinate way,” (Stalnaker, 1999).

Because dynamic semantic theories were criticized for lacking explanatory force, pragmatic theories sought to offer new solutions. For instance, in the above examples, the semantic theories have difficulties explaining how *and* is acting differently from one utterance to the next, even though it seems clear that one is acceptable and one is not. Instead of encoding information into individual lexical items (*and*, *but*, and so forth), Stalnaker suggested that perhaps the matter of updating simply came down to one of linear order.

In (20), *Moldavia is a monarchy* immediately updates the context set, eliminating all possible worlds in which Moldavia is not a monarchy. Afterwards, *the king of Moldavia is powerful* eliminates all possible worlds in which the king of Moldavia is not powerful. On the other hand, when (21) is asserted, not only does the first conjunct assert that *the king of Moldavia is powerful*, it also presupposes that Moldavia has a king, or that *Moldavia is a monarchy*. As a result, the second conjunct changes nothing about the context set, because it happens after the first. Because that conjunct violates the principle *Be Brief*, (21) is inappropriate (Schlenker, 2008).⁸

Additionally, Schlenker, in developing his theory of presupposition projection, was intent upon combining the formalism of semantics with pragmatic intuitions. To him, the content of a presuppositional phrase was dense; presupposition triggers contained a wealth of information, some of which assumed that the speaker held the presuppositional phrase to be true, as well as that the presuppositional phrase *was* true. This is in keeping with Gricean’s Maxims; namely, the Maxim of Quality (Grice, 1975).

Because these accounts view presupposition as a pragmatic phenomenon, considered through the interactions of different maxims, this means that presuppositions only surface after pragmatic processing has come into play (after a sentence’s “literal” semantic meaning has been computed) (Chemla and Bott, 2010). In practice, this means that linguistic operators apply in the standard manner, at the semantic level. In a negated assertion, the *local* interpretation will be derived first, since that is equivalent to the literal meaning of the sentence (Chemla and Bott, 2010). That isn’t to say, of course, that a global interpretation is not possible. However, it would be computed afterwards as a fall-back strategy, overruling the local interpretation.

⁸Consider the following:

- (1) It’s weird that Moldavia is a monarchy, but it is.
(Larry Horn, p.c.)

The first clause presupposes “Moldavia is a monarchy,” and since the second clause asserts it, (1) also violates *Be Brief*. So, why is (21) unacceptable, where (1) is? Horn suggests that this is the case because the second clause is “argumentatively distinct,” or opposed, to the first, and so it is not *really* a redundant affirmation (Horn, 1991).

4. Of Gaps and Gluts

4.1. The king of France, once more

- (7) The king of France is bald.

So, how *should* we think about (7)? Although Russell claimed that (7) is plainly false, Strawson asserts that “we simply *fail* to say anything true or false” (Strawson, 1950). Strawson, in other words, was putting forth an argument in favor of a non-bivalent framework, in favor of the existence of a truth-value gap. Additionally, Stalnaker, in defense of things which are undefined, says that “a sentence may fail to express a proposition at all in some possible situation,” or “it may succeed in expressing a proposition, but express one that is a partial function – one that is undefined for certain possible worlds” (Stalnaker, 1999).

Today, however, it seems as if many theories are determined to explain away a truth-value gap, instead of studying the implications of such a gap (Fodor, 1979). Certainly, a classical bivalent semantics is simpler, and preferable, but there *does* seem to be a substantial amount of evidence for *neither true nor false* which is difficult to explain away. From a semantic perspective, we have Heim’s examples marked with a #, and from a pragmatic perspective, we have what Strawson calls a feeling of “squeamishness” (Strawson (1964), Heim (1983), Von Fintel (2004)).

4.2. A further examination of gluts

- (22) Given: p is a color midway between clear red and clear orange
- p is red.
 - p is not red.
 - p is orange.
- (23) p is red or p is not red.
- p is or isn’t red.
 - p is either red, or it isn’t.
 - p Either p is red, or it isn’t.
- ⋮
- (Smith, 2008)

Examples (22) and (23) are taken from a set in Smith (2008) for which we don’t have clear ordinary speaker judgments.⁹ Can p be red and not red? Some would say yes. A less popular theory than one which includes truth-value gaps is one which include truth-value *gluts*, where utterances can be *both true and false*, or where an utterance and its negation can both be taken to be *true* (Priest, 1987). As Terence Parsons points out, gluts exist in all the same places as gaps do; these non-classical theories are truth-conditionally equivalent, and yet glut-theories are met with significantly more skepticism (and squeamishness) than gap-theories (Parsons, 1990).

⁹For further discussion, please refer to Ripley (201xa).

4.3. Bullshit?

(24) A: That man in Seat 13 is giving today's lecture.

B:

- a. # False
- b. # No¹⁰
- c. Bullshit. He's in Seat 14

(25) A: Pat's a bachelor.

B:

- a. # False
- b. # No¹¹
- c. Bullshit. He's only 10

(Ward, 2003)

Regardless of whether you buy into gaps or gluts, it seems as if these theories exist because some utterances can be made with non-catastrophic presupposition failure, with truth-conditions that do not seem to be precisely *true* or *false* (Yablo, 2006). As Gregory Ward points out, rejecting the whole of the assertion, or determining the assertion to be *false*, do not seem to be the right move, which is why he suggests *Bullshit* to be the appropriate response.¹²

Of course, when considering whether something is *true*, *false*, *neither true nor false*, or *both true and false*, we must consider whether speakers define these terms in the same way as linguists. Perhaps, where a speaker would utilize pragmatic principles to cry *Bullshit*, a semanticist, might still argue for *False*. As Janet Fodor says,

It might be said that English speakers do in fact use the words *true* and *false* as contradictories and acknowledge no third alternative. This seems to be just factually wrong, and has been admitted as such even by proponents of a two-valued logic. So the argument must be that English speakers do not use the words *true* or *false* consistently, and that they could not do so because there is no valid basis for the distinction within nontrue sentences which English speakers take themselves to be making by their restriction of the word *false* to only some nontrue sentences.

(Fodor (1979): 223)

Consequently, it seems clear that we should not dismiss gaps or gluts so easily in our theories, even if we are not sure which non-classical theory to buy into. We should also take care to differentiate between rejection of utterances because they are inappropriate (which evokes the feeling of "squeamishness") and because they are false.¹³ These distinctions will prove crucial in designing experimental paradigms.

¹⁰Unless the *No* is uttered with rising intonation and no intonational break between it and the following clause (Yablo, 2006).

¹¹Same as Footnote 5.

¹²For a more systematic treatment of denial particles, see Smith et al. (2013)

¹³As well as further examining what it means to feel "squeamish."

5. Experimental Work

In recent years, there has been a shift from relying on introspective judgments alone to gathering data from naïve speaker judgments. Though Thomason suggests that “direct reflection is obviously of little use, since Russell and Strawson seem to have honestly had opposite intuitions in this case,” we want our theory of presupposition projection to have predictions which correlate with experimental data (Thomason, 1990).¹⁴ We could say that perhaps naïve speakers do not understand *true* or *false* in the same way that linguists do, but it does seem as if naïve speaker intuitions can either corroborate our theory or suggest areas where our theory might need modifications. Discussed below are several different experimental attempts to collect data about truth-value intuitions, using methods both familiar and new.¹⁵

5.1. Presupposition projection processing

In Chemla and Bott (2010), Chemla and Bott were interested in the “projection problem” of presuppositions, outlined above. The crux of the projection problem is understanding the relationship between the presuppositions of the elementary parts of a sentence and the complex sentence itself. If they could discover exactly *where* in a sentence we process presuppositions, they thought that they could understand *how* this process of presupposition processing works.

To that end, they designed a truth-verification task, asking subjects to judge whether sentences were *TRUE* or *FALSE*. When a sentence contains a trigger for a false presupposition, competing semantic and pragmatic theories outline different predictions, both for the truth-value judgment *and* for what that means in terms of a cognitive processing model. Consider the following:

(26) Zoologists realize that elephants are birds.

(27) Zoologists do not realize that elephants are birds.

Realize is a factive verb, triggering the presupposition *elephants are birds*. This is, of course, a false presupposition. Now the question becomes, are (26) and (27) true or false?

¹⁴To be fair, Strawson’s viewpoints changed, between Strawson (1950) to Strawson (1964).

¹⁵Additionally, in *Vindication of the Humpty Dumpty Attitude Towards Language*, Herman Tennessen says this: “Of about 1,500 informants tested in some recent experiments no one seemed to act in accordance with Strawson’s predictions” (Tennessen, 1960).

In Footnote 5, he says, “I.e. almost all reacted exactly as they did to all other puzzling and ambiguous sentences: ‘If what you mean is this..., then I agree; if what you mean is..., then I disagree.’ No one would maintain that: *‘the truth or falsity question does not arise because there is no such a person as the present king of France’* (Tennessen, 1960).

Based on preliminary research, I have been unable to ascertain which prior study Tennessen is referencing. Regardless, there is a difference between the truth or falsity question not arising and an inability to answer the question. While one might maintain that the question does not arise, one might acknowledge the lack of an answer. In the framing of these experiments, it seems apparent that Tennessen, like Chemla and Bott, did not allow room for non-bivalence (Stigen, 1960).

(26) seems like it is false, but what about (27)?¹⁶ The dynamic semantics framework (Heim (1983)) would suggest that assertion (27) is false. Under this framework, negation is interpreted as the following (restated from above):

- (19) Negation [X] = Presupposition of X *and* not [Assertion of X]
(Chemla and Bott, 2010)

As we can see, the dynamic semantics model aligns with a processing model where presuppositions are interpreted *globally* first, escaping negation and processed at the level of the entire sentence. The pragmatic framework, on the other hand, suggests that assertion (27) is true:

- (28) Negation [Z realize p] = *not* [p *and* Z believe p]
(Chemla and Bott, 2010)

Cognitively, the pragmatics model aligns with the *local-first* model, where the presupposition is interpreted when it is triggered, and the negation scopes over the entirety of the sentence. Crucially, a local interpretation is also available for the global-first processing model, and vice versa, if the listener, under the Gricean Cooperative Principle, attempts to find a more charitable interpretation of assertion (27). Because both interpretations are available for both processing models, Chemla and Bott (2010) measured reaction time. If we process presuppositions in a global-first manner but choose the local interpretation, the processing times should be shorter for global interpretations (choosing *FALSE*) than local interpretations (choosing *TRUE*), whereas if we process presuppositions in a local-first manner, then the processing times should be shorter for local interpretations than for global interpretations.

Results from two experiments indicated that “participants were faster to derive a global, rather than a local, accommodation interpretation for negative sentences with a presupposition” (Chemla and Bott, 2010). Participants tended to judge assertions such as (27) as being *FALSE*, and their reaction times were faster for *FALSE* judgments than for *TRUE* ones. Chemla and Bott concluded that their results “require that a complete theory of presupposition projection include some process which explains why global interpretations arise faster than local interpretations...dynamic semantics and more generally so-called cancellation theories have a readily available explanation: the global interpretation is the root interpretation of the sentence” (Chemla and Bott, 2010).

However, although Chemla and Bott drew interesting conclusions from their data, their underlying motivation for this truth-verification task seems to have a number of potential flaws. As previously discussed, there is clearly a certain amount of semantic and pragmatic motivation to suggest a truth-value gap (or glut), something that is possibly neither true nor false, or perhaps both true and false.

Additionally, in Chemla and Bott’s second experiment, participants were given the following post-experiment questionnaire:

¹⁶The difference in intuitions for (26) and (27) is largely due to the fact that (26) entails that *elephants are birds*, which is false.

Does this sentence correctly describe the world in the story, or does it incorrectly describe the world in the story? Please read each of the options below and tick the option that best describes your answer. Please tick only one box.

Zoologists do not know that elephants are birds.

1. Incorrectly describes the world. The statement doesn't make sense.
2. Incorrectly describes the world. Elephants are not birds so the statement can't be true. It doesn't matter whether zoologists know it or not.
3. Correctly describes the world. Zoologists don't know that elephants are birds because elephants are not birds. Zoologists don't know things that aren't true.
4. None of the above. Please write down whether you think the sentence is correct or incorrect and explain your reason.
(Chemla and Bott, 2010)

The objective of this post-experiment questionnaire was to eliminate participants who did not have access to *TRUE* or *FALSE* judgments. Seven participants chose (1), while two participants chose (4) (Chemla and Bott, 2010). Although those numbers do not seem very high, only forty-five students were tested, meaning that 20% of these students rejected a particular truth-value for the statement. Moreover, the questionnaire was still framed in terms of a binary truth-value system, and it's not clear that the respondent distribution would have been the same if there had been an option of "neither true nor false" or "both true and false."

5.2. Experimenting with the king of France

Although Chemla and Bott were working within a binary truth-value system, others have been exploring alternative systems, particularly in regards to classic examples such as these:

- (7) The king of France is bald.
- (29) The exhibition was visited yesterday by the king of France.
- (30) The king of France is sitting in that chair.

Although intuitions are mixed with regards to (7), assertions (29) and (30) appear to be definitively false (Strawson (1964), Fodor (1979)).¹⁷ The king of France does not exist, but there is somehow something different about asserting his baldness, as opposed to an action he is taking. Various factors, such as topic-comment structures or background knowledge have been identified as possible influences on our truth-value intuitions about sentences where the noun phrases fail to refer (Abrusán and Szendrői (2012), Strawson (1964), Fodor (1979), Von Fintel (2004)). In (Abrusán and Szendrői, 2012), Abrusán and Szendrői investigated the truth-value intuitions of these kinds of sentences, examining the influence of verifiability and topicality.

¹⁷Indeed, (30) can be compared to Yablo's (10), which is also considered to be a false statement.

Participants were presented with experimental items like the following:

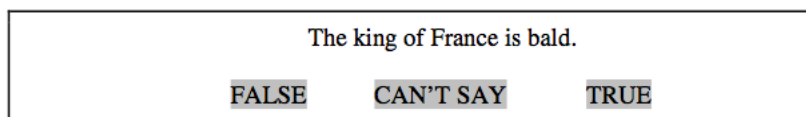


Figure 1: Experimental item: Abrusán and Szendrői (2012)

Although the inclusion of *TRUE* and *FALSE* are self-explanatory, the inclusion of *CAN'T SAY* is what distinguishes this experimental paradigm from that of Chemla and Bott (2010). The assumption, here, was that “if participants felt ‘squeamish’ about assigning a truth-value to a particular item, they would press the *CAN'T SAY* label” (Abrusán and Szendrői, 2012).¹⁸

Notably, participants responded with *FALSE* to all of the positive conditions, while the negated versions of the same items were not judged as uniformly *TRUE*, producing an asymmetry in the judgments of positive and negative conditions, despite the fact that presuppositions do not change when assertions are negated. If we believe the Russellian story, because the positive sentence entails the existential proposition “The king of France exists” and the negative sentence does not, this could explain the asymmetry. However, Abrusán and Szendrői (2012) concluded that their “results thus support the idea in Lasersohn (1993) and von Fintel (2004) according to which semantic truth-value and the intuitive truth-value felt by speakers are not the same things.” (Abrusán and Szendrői, 2012). Where Chemla and Bott (2010) compared semantic and pragmatic frameworks strictly in terms of *True* and *False*, Abrusán and Szendrői (2012), based on von Fintel’s work, claim that

our intuitions about accepting or rejecting a sentence as true or false, and the sentence’s actual semantic truth-value (and hence its presuppositionality) are two separate things. Speakers might feel that a sentence is false or true even when semantically it has no truth-value, as long as they can find some reason based on which they can reject (or accept) the sentence. The feeling of ‘squeamishness’ arises only when all pragmatic repair strategies for dealing with a truth-valueless sentence fail.

(Abrusán and Szendrői (2012): 4)

18

- (1) The Yankees will win their next game
- (2) It will rain a week from tomorrow”

Larry Horn, in personal correspondence, noted that an experiment involving future contingents, such as (1) or (2), could be of interest, if presented with these three response choices. I agree with his prediction of a large number of *CAN'T SAY* responses. Moreover, the motivation to choose *CAN'T SAY* is not necessarily correlated with an intuitively nonbivalence framework; it is far more likely that participants feel that they have insufficient information. For example, we could say that (1) *does* have a truth-value; we just don’t know it yet. Essentially, all Abrusán and Szendrői (2012) did was test for the presence of “squeamishness,” as well as what contributes to that feeling, though there appear to be a number of reasons why we might feel this way.

For the purposes of Experiments 1-3 (see below), the *kind* of truth-value of interest is what Abrusán and Szendrői (2012) call the *felt truth-values*, where the truth-value gap can indeed be characterized as a feeling of “squeamishness.”

5.3. Fat Tony might be dead

- (31) *Given:* Fat Tony, a mobster, has faked his own death.
 Expert A: Fat Tony is dead.
 Expert B: Fat Tony might be dead.

Expert A’s assertion is a straightforward assertion, while Expert B’s assertion contains an epistemic possibility modal. What is the truth-value of Expert B’s utterance?

Seth Yalcin and Joshua Knobe could have designed their experiment in a number of ways. Like Chemla & Bott (2010), Yalcin and Knobe could have simply given participants two options: *True* and *False*. Or, like Abrusán and Szendrői (2012), Yalcin and Knobe could have given participants three options, if they wanted to explore non-bivalent theories: *False*, *Can’t Say*, and *True*. However, participants in Yalcin and Knobe (2010) were randomly presented with the short story, and one of the following:

Table 1: Conditions

Condition	Statement
(IS-True)	What Expert A said is true.
(IS-False)	What Expert A said is false.
(MIGHT-True)	What Expert B said is true.
(MIGHT-False)	What Expert B said is false.

The results for the IS-groups were as expected (because Expert A’s utterance is so straightforward). Participants disagreed strongly in the IS-TRUE group, but agreed strongly in the IS-FALSE group. In the MIGHT-group, however, although participants strongly rejected “What Expert B said is false,” this rejection was not matched by a strong tendency to *agree* with “What Expert B said is true.” Indeed, as Yalcin & Knobe (2012) report, “on balance, then it appears that subjects were divided between (a) the view that what Expert B said is true and (b) the view that what Expert B said is neither true nor false.” Notably, these results would not necessarily have been the same if participants had been presented with a “neither true nor false” option. For one, although ordinary subjects are familiar with questions of ordinary truth and falsity, the concept of something being “neither true nor false” is familiar primarily to semanticists and pragmaticists, and this particular experimental design is one way of eliciting such judgments, which produced the basis of the design of Experiment 1.

Additionally, instead of the traditional method of recruiting college undergraduates, participants were recruited via an online platform, known as Amazon Mechanical Turk. The advantages and disadvantages of this platform are discussed below.

6. Experimental Design

6.1. What is Amazon Mechanical Turk?

Amazon Mechanical Turk was originally developed as an in-house tool because Amazon recognized that there were some tasks which humans performed with more efficiency and accuracy than computers. Its name is derived from the story of “The Turk,” a chess-playing automaton constructed in 1770 by Wolfgang von Kempelen to impress the Empress Maria Theresa of Austria (Atkinson, 1998). The automaton appeared to play a highly skilled game of chess, but it was later revealed that the machine was operated by a chess master hidden inside.

Today, MTurk (as it is colloquially known) is one of the largest online labor markets, with over 100,000 users from over 100 countries, completing tens of thousands of tasks daily (Pontin 2007). Employers, known as *Requesters*, submit *HITs*, or Human Intelligence Tasks, for *workers* to complete. While MTurk began as a labor market for tasks such as picture identification or data entry, it is now used as a subject pool for experimental research, most often in psychology and market research. On MTurk, running hundreds of participants within a few days is easy and inexpensive, making it attractive as a subject pool.

Because MTurk is a relatively new service, and because there are often preconceived notions about the demographics of internet users, there may be some skepticism about the validity of research conducted via MTurk. To that end, the demographics of American MTurk *workers* are discussed below. For the purposes of the current research, all non-American *workers* were excluded via IP address tracing, since we were only looking for subjects whose native language was American English.

6.1.1. The demographics

Currently, the majority of studies are conducted on undergraduates, a subject pool which is not necessarily representative of the larger population. For one, undergraduates are, for the most part, WEIRD: Western, educated, industrialized, rich, and democratic (Buhrmester et al., 2011). MTurk *workers*, on the other hand, are “slightly more representative of the U.S. population than are standard internet samples,” as well as being, more importantly, “significantly more diverse than typical American college samples” (Buhrmester et al., 2011). In multiple studies, MTurk *workers* have been shown to be diverse in age (20-70 years), educational background (holders of bachelor’s degrees are overrepresented), and socioeconomic status (with household incomes ranging from less than \$10K to over \$300K) (Gosling et al. (2004), Behrend et al. (2011), Berinsky et al. (2011), Ipeirotis (2010)). Additionally, if a *worker* has completed many HITs with a high level of accuracy, they may qualify to be a *Master worker* (Behrend et al., 2011).

For a more in-depth discussion of the demographics of MTurk *workers*, please see *Appendix A.1*.

6.1.2. Participant compensation

Aside from skepticism about the MTurk demographics, questions are often raised about subject compensation. In linguistics research using surveys, for instance, subjects are typically compensated with \$5.00- \$10.00 (or course credit) for 20-60 minute surveys (depending on the policies of the university and/or the Institutional Review Board). On MTurk, the market can fluctuate, but *workers* are usually paid \$0.02 for a 5 second HIT, and Amazon recommends compensating \$5.00-6.00 an hour.

Although this may appear to be a low amount to pay for labor, self-described *Turkers* are not necessarily completing HITs for the monetary reward. For the purposes of the current research, most American *Turkers* report that they spend time on MTurk “to kill time,” because they “find the tasks to be fun,” or as a source of secondary income for “gadgets and hobbies” (Ipeirotis, 2010).

6.1.3. Participant recruitment and data quality

Although compensation is not a primary factor, it is still an important factor to consider when pricing a HIT. “Participation is affected by compensation rate and task length but participants can still be recruited rapidly and inexpensively” (Buhrmester et al., 2011). While Buhrmeister et al. were able to recruit participants for a 30 minute survey for as little as \$0.02, they did not exclude any potential participants, and it appears as if American *workers* demand higher compensation than *workers* from other countries (which, given that the standard of living in America is higher, makes sense). When only allowing for American *Master workers*, I was unable to recruit subjects for a 20 minute survey by paying less than \$1.00, likely because *Master workers* demand higher compensation than other *workers*. Recruiting enough participants also took significantly more time, whereas relaxing the qualifications to completion of a minimum of 1000 HITs with a 95% accuracy rating resulted in faster recruitment of subjects (100+ within 5 hours).

A common concern about anonymous participants recruited online is about data quality, that participants might not take the research seriously, choosing to complete a survey randomly in order to earn the payment. However, on MTurk, it is possible for a *requester* refuse payment for subpar work, thus lowering a *worker’s* approval rating and reducing the number of HITs they will be able to see and complete in the future. In addition, “realistic compensation rates do not affect data quality” (Buhrmester et al., 2011).

6.1.4. Using MTurk for experimental research

The concerns about using MTurk for experimental reasons include concerns about differences in online/offline participant demographics, compensation, and data quality. However, prior research shows that there are not significant differences in the demographics, the lowered rates of compensation do not affect data quality, and there are quality control measures in place to prevent participants from purposefully performing poorly. “The data obtained are at least as reliable as those obtained via traditional methods (Buhrmester

et al., 2011). As a result, “the use of these labor portals [such as Amazon Mechanical Turk] is an efficient and appropriate alternative to a university participant pool” (Behrend et al., 2011).

6.2. Stimuli

Chemla and Bott have this to say on the formulation of their stimuli:

Each sentence was formed using *Geographers* or *Zoologists* as the subject, a factive verb with negation (*do not realize*) or a non-factive verb (*tell*), and a proposition about categories as the complement of the verb (e.g., *elephants are mammals* or *elephants are birds*).

We generated 60 place names and 60 animals as the subordinate category member of the category proposition. These exemplars formed the basis of each item in the design. The experimental sentences were generated using an exemplar and an incorrect super category, combined with the appropriate professional (zoologists or geographers)...Four control versions of each item were formed using the same exemplar but with a different subject or a different superordinate category (i.e. the correct superordinate) to obtain unambiguously true and false sentences, with the presuppositional phrase (*do not realize* and without a presuppositional phrase (*were told*).

No subordinate category member were used more than once but superordinate category members were used multiple times. There were four superordinate geographical categories (Africa, Asia, Europe, America) and six superordinate zoological categories (birds, dogs, fish, insects, mammals, and reptiles). Each superordinate appeared equally across conditions.

Items were assigned to five counterbalancing lists (distributed equally among participants) so that all items appeared equally often in each condition, but no participant saw the same exemplar twice.

(Chemla and Bott (2010): 7-8)

Items were assigned to ten counterbalanced lists (distributed equally among participants) so that all items appeared equally often in each condition, but no participant saw the same exemplar twice. Although participants saw 24 items in each condition in Chemla and Bott’s study, participants via MTurk saw 6 items in each condition for a total of 30 items. The decision to cut down on the stimuli was due to the fact that online participants tend to have a shorter attention span (Buhrmester et al., 2011).¹⁹ See below in Table 2 for examples of test sentences and controls (adapted from Chemla and Bott (2010)).

¹⁹Please see Appendix C.3 for a full list of the stimuli used.

Table 2: Example Sentences

Condition	Example sentences	Expected answers	“True?”	“False?”
(a)	Zoologists do not realize that elephants are reptiles.	True/False	Yes/No	Yes/No
(b)	Zoologists do not realize that elephants are mammals.	False	No	Yes
(c)	Geographers do not realize that elephants are mammals.	True	Yes	No
(d)	Zoologists were told that elephants are mammals.	True	Yes	No
(e)	Zoologists were told that elephants are reptiles.	False	No	No

7. Experiment 1

7.1. Participants

150 MTurk *workers* participated in a ten minute Qualtrics survey for \$0.50. 29 participants were removed for poor performance (<0.75 accuracy) on the control sentences. All participants were required to have completed at least 1000 HITs with a 95% accuracy rating, to control for data quality. All *workers* self-identified as native English speakers, and all IP addresses were tied to the U.S. Participants ranged in age from 18-66 (with a mean of 35.4 and a standard deviation of 12.3). 77 males and 73 females participated.

7.2. Design

For Chemla and Bott (2010), each item was presented as the following, with a *True* or *False*:

- (32) Zoologists do not realize that elephants are birds.
 True False

Stimuli for this experiment were adapted so that each participant only saw all items in the *False* or *True* group, in an experimental paradigm similar to (Yalcin and Knobe, 2010). Instead of being asked whether an item was true or false, participants were asked whether they agreed or disagreed with the assessment of the item as true or false:

- (33) Zoologists do not realize that elephants are birds.
 Is this underlined statement true/false?
 YES NO

For the experimental items, if participants agreed with “Is this underlined statement true/false?” in both the TRUE- and FALSE-groups, then they would seem to be judging those items as *both true and false*. If participants disagreed with that statement in both the TRUE- and FALSE-groups, then they would seem to be judging those items as *neither true nor false*. If they agreed with that statement in the TRUE-group but not in the FALSE-group, then they would seem to be judging those items as *true*, and if they agreed with that statement in the FALSE-group but not in the TRUE-group, then they would seem to be judging those items as *false*.

7.3. Procedure

Participants were given a cover story to remove any potential ambiguity about the situation, allowing for the construction of unambiguously true and false control sentences, with and without presupposition triggers. After the cover story, participants went through a training phase in which they judged 16 control sentences (4 from each control condition) and received feedback on their responses. No experimental sentences were presented during the training phase and participants did not receive feedback during the main part of the experiment.

7.4. Results

7.4.1. Data treatment

75 MTurk *workers* participated in the TRUE-group and 75 MTurk *workers* participated in the FALSE-group, for a total of 150 participants. All participants received payment (\$0.50) for their participation. 29 of the participants were removed for poor performance on the control conditions (scoring with less than 75% accuracy).

7.4.2. Comparing to Chemla and Bott’s Results

Note: Condition (a) is the experimental condition. Conditions (b-e) are control conditions. Standard deviations are in parentheses.

Table 3: Conditions and Observed Accuracy
(Chemla and Bott, 2010)

Condition	Expected answers	Observed “true” proportions
(a)	True/False	0.38 (0.32)
(b)	False	0.12 (0.073)
(c)	True	0.85 (0.094)
(d)	True	0.93 (0.052)
(e)	False	0.11 (0.079)

Table 4: Exp. 1: Conditions and Observed Accuracy
TRUE-Condition

Condition	Expected answers to “Is it true?”	Observed “yes” proportions
(a)	YES/NO	0.29 (0.40)
(b)	NO	0.03 (0.10)
(c)	YES	0.94 (0.16)
(d)	YES	0.97 (0.08)
(e)	NO	0.06 (0.11)

Chemla and Bott (2010)’s results of their participants’ choice proportions are from 0 to 1, where 0 indicates that no participants judged any items in that condition to be true,

Table 5: Exp. 1: Conditions and Observed Accuracy
FALSE-Condition

Condition	Expected answers to “Is it false?”	Observed “yes” proportions
(a)	YES/NO	0.64 (0.31)
(b)	YES	0.85 (0.21)
(c)	NO	0.33 (0.36)
(d)	NO	0.22 (0.23)
(e)	YES	0.69 (0.28)

and 1 indicates that all participants judged all items in that condition to be true. On the control conditions, participants performed fairly well, and in the experimental condition, participants judged items to be false more often than true.

Like the participants in Chemla and Bott (2010), MTurk *workers* in Experiment 1 performed well on the controls, demonstrating an understanding of the cover story.²⁰ In addition, in the TRUE-group, the majority of participants rejected the experimental items as *true*, whereas in the FALSE-group, although participants were more likely to accept the experimental items as *false*, they were more evenly split.²¹ However, it *was* statistically significant ($p < 0.05$) that participants accepted the premise that the experimental items were false.²²

7.5. Discussion

From the data, we can see that, although participants had few qualms about judging experimental items as *not-true*, they were split as to whether items were *false* or *not-false*. Histograms of participant responses, however, reveal more drastic differences in participant behavior:

Although we might expect these histograms to be mirror images of each other, participants clearly pattern differently in each group. In the TRUE-group, 30+ participants (almost half) judged all of the experimental items as *not true*, which indicates that participants were fairly consistent in their judgments. In the FALSE-group, however, participants were much less consistent; only 13 participants judged all experimental items as *false*.

Additionally, although Chemla and Bott (2010) show comparable “true” proportions between Conditions (b) and (e), the two conditions for which the expected answer is *FALSE*, participants in Experiment 1, in the TRUE-group, did not perform as well. In the FALSE-group, they did significantly worse. Below are examples of Conditions (b) and (e), respectively:

²⁰Participants also performed better in the TRUE-group than in the FALSE-group

²¹If participants had been definitively split on the experimental items, the expected mean would be a 0.5, but since the mean was a 0.64, we know that participants were slightly more likely to agree that the experimental items were false.

²² $p = 0.001$

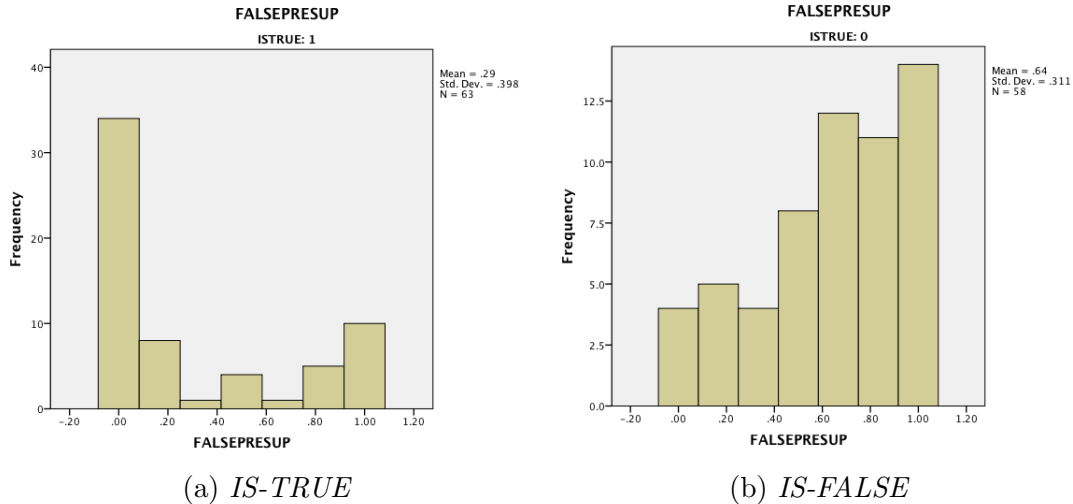


Figure 2: Condition (a) Histograms

- (34) Condition B:
Geographers do not realize that London is in Europe.
- (35) Condition E:
Geographers were told that London is in America.

In pilot studies, and in Experiments 2 and 3, this difficulty never arose, and so the origin of this anomaly seems unclear. Perhaps participants were simply mistaken about some basic facts about the world. For instance, nearly all participants judged (36) inaccurately:²³

- (36) Condition E:
Zoologists were told that doves are mammals.

Alternatively, there may have been some difficulty with the premise of the cover story. Nonetheless, most participants still scored >0.75 on the fillers, demonstrating an understanding of the story. But, given that participants wavered so much between *false* and *not false*, this raises a few questions for Chemla & Bott’s conclusions about the cognitive processing model, since those conclusions were drawn from a reaction time experiment based on these same stimuli (Chemla and Bott, 2010).

8. Experiment 2

8.1. Participants

150 MTurk *workers* participated in a five-minute Qualtrics survey for payment (\$0.50). 47 of the participants were removed for poor performance on the control conditions (scoring with less than .75 accuracy). All participants were required to have completed at

²³For the record, doves are birds, and birds are not mammals. In case there was some uncertainty about the matter.

least 1000 HITs with a 95% accuracy rating, to control for data quality. All *workers* self-identified as native English speakers, and all IP addresses were tied to the U.S. Participants ranged in age from 18-71 (with a mean of 33 and a standard deviation of 10.5). 57 males and 93 females participated.

8.2. Design

Because the design of Experiment 1 asked participants to choose between *YES* and *NO*, the experiment was still displayed to participants in a binary fashion. As a result, it is conceivable that participants were still operating within a forced bivalent framework, believing that if they rejected whether an item was true, it was false, and vice versa, without considering any alternatives. Because it was not statistically significant that participants chose *not false*, perhaps the indirect method was not the best way to test for a third value. If a third option was actually offered, perhaps participants would feel more confident in their choices.

Experiment 2, designed with the same answer choices as Abrusán and Szendrői (2012), offered *CAN'T SAY* as an alternative to *FALSE* or *TRUE*. *CAN'T SAY* is suggested to be the choice for when a participant feels “squeamish” about the truth-value of an item (Abrusán and Szendrői, 2012). An example item is given below:

(37) Zoologists do not realize that elephants are birds.

FALSE CAN'T SAY TRUE

8.3. Procedure

The procedure was similar to Experiment 1, excepting that participants did not receive feedback for all of the training sentences (which would likely have lead to a participant bias against the *CAN'T SAY* option). An example of an experimental item on Qualtrics is seen below:



Figure 3: Condition (a)

Instructions at the top of each page duplicated Abrusán and Szendrői (2012), making it explicit when a participant should click *CAN'T SAY*:

(38) *In this experiment, statements will appear on your screen. If you think a statement is true, you should click on the 'TRUE' button. If you think a statement is false, you should click on the 'FALSE' button. Sometimes, it may happen that you cannot decide. In those cases, you should click on the 'CAN'T SAY' button.*

Please do not dwell on your decision for too long. There is no right or wrong answer!

8.4. Results

8.4.1. Data treatment

150 MTurk *workers* participated for payment (\$0.50). 47 of the participants were removed for poor performance on the control conditions (scoring with less than 75% accuracy). All *workers* self-identified as native English speakers, and all IP addresses were tied to the U.S.

8.4.2. Choice Proportions

Note: Conditions (b-e) are control conditions. Standard deviations are in parentheses.

Table 6: Exp. 2: Control Conditions and Observed Accuracy

Condition	Expected answers	Observed “correct” proportions
(b)	False	0.95 (0.10)
(c)	True	0.92 (0.13)
(d)	True	0.99 (0.52)
(e)	False	0.76 (0.26)

As in Experiment 1, participants did well across Conditions (b-d) (all of which were controls), but Condition (e) proved difficult. Items in Conditions (d) and (e) are reproduced below:²⁴

- (39) Condition D:
Geographers were told that Florida is in America.
- (40) Condition E:
Geographers were told that Florida is in Asia.

If we examine participant responses more closely, we see that although participants were fairly unanimous for conditions (b-d), they occasionally opted for *CAN’T SAY* in condition (e), and were pretty divided in the experimental condition (condition (a)):

Exact percentages for participant responses for condition (a) are given below:

Additionally, in a paired-samples t-test, *CAN’T SAY* was statistically significant as compared to Conditions (b-d) ($p = 0.001$), but not as compared to Condition (e) ($p = 0.025$).

²⁴Items in Condition (e) are false. For example: Florida is not in Asia, and the geographers were only told true information about geography. Although we could hypothesize that perhaps participants did not fully understand the cover story, they had no difficulty with Condition (d), so we are left to be perplexed about their knowledge of geography.

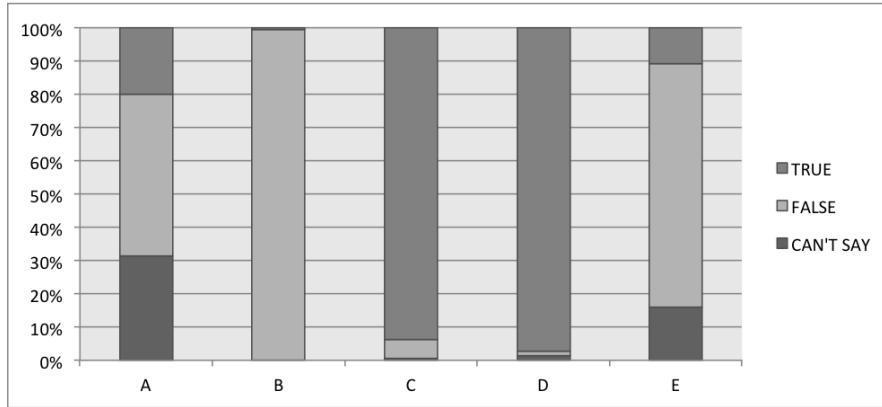


Figure 4: Responses by Condition

Table 7: Exp. 3: Condition (A) Responses

Response	Percentage
False	0.49
Can't Say	0.30
True	0.21

8.5. Discussion

To a certain extent, it was expected that a large number of participants in Experiment 1 would not perform well on the controls, due to the fact that the experimental paradigm (particularly in the FALSE-group) was a fairly novel one, and participants might have been unfamiliar with it. Aside from adding a *CAN'T SAY* response choice, which might have been unfamiliar, the design of Experiment 2 was a fairly standard one, which is why it is surprising that almost one third of the participants in Experiment 2 performed poorly on the control conditions. It is likely that this is due to the fact that participants did not fully comprehend the premise of the cover story.

Additionally, there was little feedback in the training portion. One of the flaws with a study conducted online is that participants are unable to ask the experimenter regarding any questions which they might have, and it is possible that more participants would have performed more accurately if the task was conducted offline, with the experimenter in the room. And not only did participants lean towards judging items in the experimental condition as *FALSE*, their choice of *CAN'T SAY* was not statistically significant ($p = 0.102$).

One possibility is that participants were biased against choosing *CAN'T SAY*, for fear of appearing “stupid.” Participants might have felt that choosing *CAN'T SAY* indicated their inability to determine an item’s truth-value because they were too unintelligent to do so, or because they felt that they had insufficient information, as opposed to determining that an item did not *have* a truth-value. Abrusán and Szendrői (2012) state that they “made sure that avoidance of the ‘CAN’T SAY’ label was not due to a general unwillingness on the part of [their] participants to admit their ignorance.” Again, however, there may be a difference between participant biases on- and offline, and I don’t believe

that we should rule out the “fear of looking stupid” as a participant bias.

9. Experiment 3

9.1. Participants

75 MTurk *workers* participated for payment (\$0.50). 8 of the participants were removed for poor performance on the control conditions (scoring with less than 0.75 accuracy). All participants were required to have completed at least 1000 HITs with a 95% accuracy rating, to control for data quality. All *workers* self-identified as native English speakers, and all IP addresses were tied to the U.S. Participants ranged in age from 18-69 (with a mean of 38.0 and a standard deviation of 13.8). 35 males and 40 females participated.

9.2. Design

Perhaps *CAN'T SAY* isn't the best representation of a truth-value gap option, particularly for naïve speakers. Moreover, in testing for gaps, why not test for gluts as well? For Experiment 3, *CAN'T SAY* was replaced with *NEITHER*, to represent the truth-value gap option, and *BOTH* was added to represent the truth-value glut option. A sample experimental item is seen below:

(41) Zoologists do not realize that elephants are birds.

FALSE TRUE NEITHER BOTH

In Experiment 2, *CAN'T SAY* was placed between *TRUE* and *FALSE*, perhaps incorrectly suggesting that *CAN'T SAY* falls somewhere between *TRUE* or *FALSE*, as opposed to falling outside of this system. In comparing indirect/direct response choice offerings, Experiments 1 and 3 are more matched than Experiments 1 and 2.

9.3. Procedure

The procedure was the same as Experiments 1 and 2, but participants did not receive feedback for all of the training sentences (which would likely have lead to a participant bias against *NEITHER* or *BOTH*, because the control conditions were definitively *TRUE* or *FALSE*). An example of an experimental item on Qualtrics is seen below:

9.4. Results

9.4.1. Data treatment

75 MTurk *workers* participated for payment (\$0.50). 8 of the participants were removed for poor performance on the control conditions (scoring with less than 0.75 accuracy).



Figure 5: Condition (a) on Qualtrics

All *workers* self-identified as native English speakers, and all IP addresses were tied to the U.S.

9.4.2. Choice Proportions

Note: Conditions (b-e) are control conditions. Standard deviations are in parentheses.

Table 8: Exp. 3: Control Conditions and Observed Accuracy

Condition	Expected answers	Observed “correct” proportions
(b)	False	0.99 (0.04)
(c)	True	0.86 (0.23)
(d)	True	0.97 (0.10)
(e)	False	0.98 (0.09)

Unlike Experiments 1 and 2, participants had no difficulties with Condition (e), although Condition (c) seemed to cause some trouble. Condition (c) is the wrong subject condition:

(42) Zoologists do not realize that Kentucky is in America.

Regardless of whether Kentucky is in America (which it is), the cover story is clear on the fact that zoologists do not know anything about biology, and biologists do not know anything about zoology. So, Condition (c) should be judged to be *true*, under this cover story (although the cover story could conflict with a participant’s real-world knowledge). Given that only 8/75 participants had to be excluded from the analysis for poor performance on the controls (as opposed to 47/150 participants in Experiment 2), it seems as if participants had a better grasp of the cover story in Experiment 3 than in Experiment 2. Experiment 3’s Conditions (c) and (d) proportions (0.86 and 0.97) are also comparable to Chemla and Bott (2010)’s results (0.85 and 0.93, respectively), where participants did worse on Condition (c) as well.

And in a paired samples t-test, choosing *NEITHER* was statistically significant ($p < 0.05$) for the experimental condition, as opposed to all of the control conditions. On the other hand, choosing *NEITHER* for the experimental condition was not statistically significant, as opposed to the other answer choices ($p = 0.315$)²⁵. It *was* statistically that participants

²⁵I conducted a one-sample t-test, comparing *NEITHER* to 2/6 (or 0.33), or to the chance that *NEITHER* had been chosen at random. If *BOTH* had been more of a viable option, then *NEITHER* would have been compared to a test value of 0.25, but the effect of *BOTH* seemed negligible at best.

chose *FALSE* ($p < 0.5$), which is similar to Experiment 1.

Table 9: Exp. 3: Experimental Items Response

Response	Percentage
FALSE	0.455
TRUE	0.259
NEITHER	0.286
BOTH	0.00

In addition, unsurprisingly, *BOTH* was never chosen as a response choice. The data from 67 participants were analyzed, for a total of 2010 responses to items. Out of all 2010 responses, *BOTH* was only chosen 6 times, or 0.3% of the time (and none of those times were in response to experimental items).

9.5. Discussion

Although participants appeared to have some difficulty with control condition (c), overall, they seemed to have a good understanding of the cover story. There was also no difficulty with control condition (e), which participants found to be more difficult in Experiments 1 and 2, though only in a statistically insignificant way. And though it does seem conceivable that this could be due to the fact that the explicit concept of *both true and false* is a novel idea to an ordinary speaker, it does not seem shocking that *BOTH* was chosen so few times.²⁶

In the experimental condition, participants judged items most often to be *FALSE*, though *NEITHER* and *TRUE* were of equal weight.

10. Conclusion

Despite the fact that Experiments 1-3 did not provide statistically significant evidence in favor of gapped theories, there are still findings of note for future experimental work. Because these experimental paradigms were new for participants, an online survey may not have been the most ideal way to conduct these experiments. Though it is efficient and inexpensive to run large groups of participants in a matter of days, if the participants cannot ask the experimenter for feedback, and feedback on response choices during the training sessions could potentially bias the participants (see Experiments 2 and 3), it seems as if experiments ought to be crowdsourced using online platforms when they are familiar (or easy) paradigms. The number of participants who had to be removed from the data analysis in Experiment 2 was close to a third, for instance.

Additionally, though *true* and *false* appear as common answer choices on standardized exams, for instance, it seems doubtful whether ordinary speakers have ever been asked to consider whether a statement is *neither true nor false* or *both true and false*, outside of

²⁶Though evidence has shown that the concept of a glut is not so difficult for vague predicates.

participating in psycholinguistic experimental work. Moreover, given that a multivalent framework is not the standard experimental paradigm, the indirect method of testing for an ordinary speaker’s intuitions seems to be more useful than any direct methods.

Even among gappy frameworks, there is little consensus as to how to define the “gap” - indeed, how can we come up with a consistent typology for “gaps,” anyway? ²⁷ We would require one which is understandable for both a linguist and an ordinary speaker. And as a side note, it does not currently seem useful to test ordinary speaker intuitions for “gluts,” at least insofar as factive verb triggers are concerned. Some work has been done to show that ordinary speakers are more willing to judge utterances such as *John is both tall and not tall* as *both true and false*, but those judgments about relatives do not appear to extend to factives Ripley (201xa), Ripley (201xb)).

It seems clear that participants reject the truth of sentences with false presuppositions triggered by factive verbs, but as to whether ordinary speakers believe these sentences to be false or feel “squeamish” about assigning them a truth-value, that is a matter that best left to future experimental work. Certainly, if we are to give any weight to any presupposition projection theories where negated utterances containing false presuppositions turn out to be *true*, we should at least give the same consideration to theories where these utterances are judged to have no truth-value, since these experiments show a preference for *neither true nor false* over *true*. But how should we theorize about these wavering “felt” truth-value judgments?

Perhaps, instead of continuing to attempt to explain away truth-value gaps, we ought to look back at Strawson, and consider these situations where the question of truth or falsity does not arise as valid and plausible (Strawson, 1950). But even if we form a theory which can accommodate gaps, we must determine when and where these gaps arise. We must account for this feeling of “squeamishness.” And as we know from much work around presuppositions, they are context-dependent; are these presuppositions in the common ground, or not? How can we verify that they are true? How is it that the same content of an utterance can be true, or false, or neither true nor false? Does it matter how the utterance is structured?²⁸ What can be assumed to be true, and what false presuppositions are we willing to overlook? As Abrusán and Szendrői (2012) pointed out, “speakers might feel that a sentence is false or true even when semantically it has no truth-value, as long as they can find some reason based on which they can reject (or accept) the sentence.” So, what exactly are these “reasons” which speakers can find?

From these kinds of questions, it seems obvious that pragmatic factors come into play, and that we should give these factors some careful consideration.

²⁷The same worry arises for “gluts.”

²⁸Topic-comment, for instance, seems to play a role, and we know that intuitions differ for “The king of France is bald” (Ex. (7)) and “The exhibition was visited by the king of France yesterday” (Ex. (29)).

10.1. Next steps

Before we can begin to draw any conclusions about a cognitive model of presupposition processing, we must first take a closer look at truth-value intuitions. What we can learn from Abrusán and Szendrői (2012) is that we need to consider the differences between semantic truth-values and “felt” truth-values, or, the truth-values we assign once we take into account pragmatic factors. Although topic-comment structures and verifiability have been tested for definite descriptions, we should extend this line of experiments to various presupposition triggers, in order to examine what *are* these factors which influence our choice between *false* and *not false* (whether *not false* is *true* or *not true* or *false*), as well as how much weight should be given to each factor. And in establishing these differences we should take into account the ways in which we design our response choices, since, given the same stimuli, presented in different ways, participants will respond in different ways.

Additionally, more thought should be given as to how internally consistent a theory of presuppositions must be. Though definite descriptions and factives have traditionally been considered to pattern together, *CAN'T SAY* appears to be a plausible response to utterances containing definite descriptions which do not refer, as opposed to utterances containing false propositions triggered by factive verbs (Abrusán and Szendrői, 2012).²⁹ Perhaps we must redefine what it means to be a “presupposition,” since we can not even come to a consensus about the various trigger classes, nor do we have good tests for the existence of a presupposition. The “Hey, wait a minute test,” used in Von Stechow (2004) and Yablo (2006), has been shown to be pretty unreliable:

- (43) A: I ate some of the cookies.
B: Hey, wait a minute (you didn't eat **SOME** of them) – you ate them all!
- (44) A: Not only is he honest and moral, he's electable.
B: Hey, wait a minute – he doesn't know the meaning of the word honest.
(Examples from Larry Horn, personal correspondence)

If we have a clearer picture of what presuppositions are, and what kinds of truth-values exist (and when and where), perhaps *then* we can formulate a theory of presupposition projection which better predicts our existing data.

²⁹It is possible that this is due to the fact that it is fairly easy to determine whether or not a person existed, and to judge an utterance accordingly, whereas facts about the world are more difficult to ascertain.

A. Appendix: MTurk and Qualtrics

Because MTurk is still a relatively new platform through which to run studies, it seems useful to include notes on various ways to do so. For additional resources, I recommend referring to the blog Experimental Turk, found here: <http://experimentalturk.wordpress.com>

A.1. MTurk Demographics: a closer look

As we can see in Figure (6), American MTurk *workers* tend to be younger, though the participants in Experiments 1-3 ranged in age from (18-75):

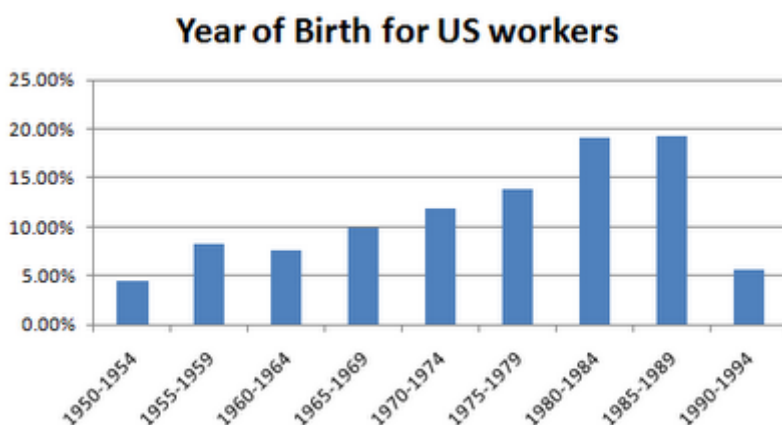


Figure 6: (Ipeirotis, 2010)

Workers are also pretty highly educated, which might be considered to be surprising, given the number of fairly menial HITs on MTurk:³⁰

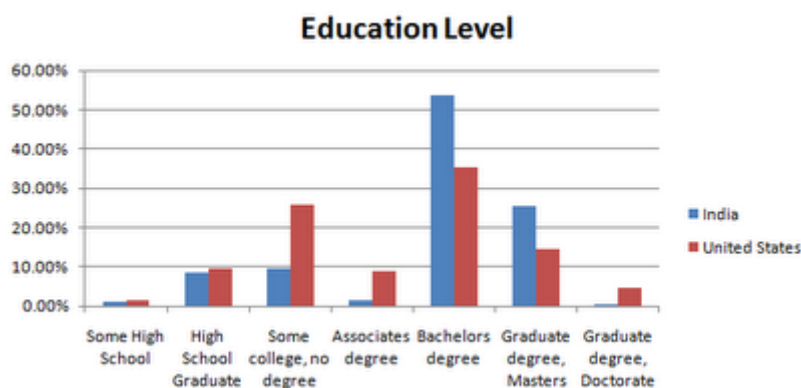


Figure 7: (Ipeirotis, 2010)

³⁰Numbers are given for India as well because on MTurk, after Americans, Indian *workers* are the second highest demographic, though *workers* in America and India use MTurk for different purposes (primary v. secondary/tertiary sources of income, etc.)(Ipeirotis, 2010).

Workers also skewed towards being female (in the U.S.), which proved to be true as well in Experiments 1-3:

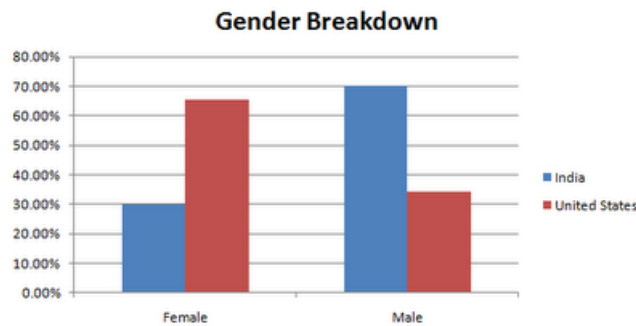


Figure 8: (Ipeirotis, 2010)

And based off the distribution of household income for American *workers* and income per week from Amazon MTurk, it seems fair to say that MTurk is not a primary source of income for American *workers*.³¹

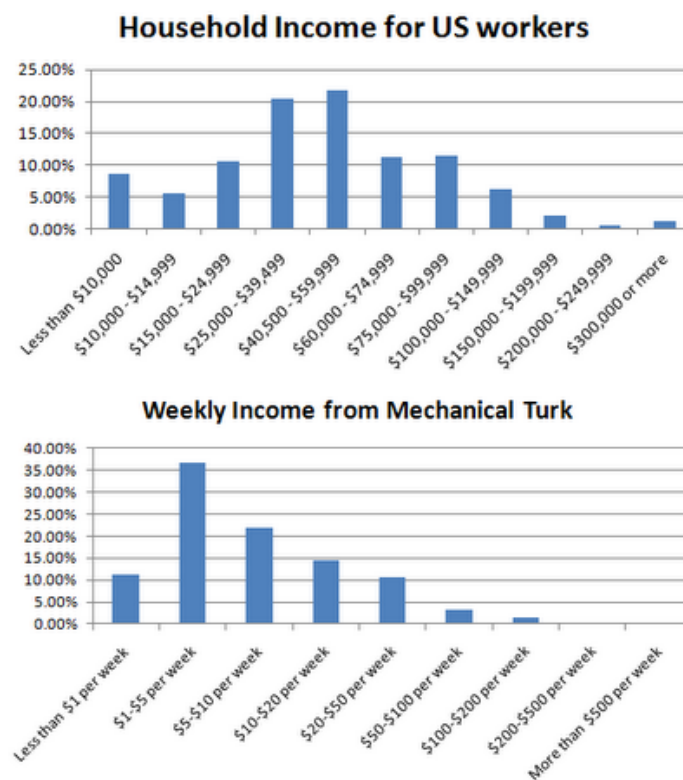


Figure 9: (Ipeirotis, 2010)

Additionally, as compared with non-MTurk participants, MTurk participants were less extraverted, less emotionally stable, and had lower self-esteem; MTurk participants also exhibited attitudes about money and time that were more similar to student participants than to community participants (Goodman et al., 2012).

³¹Thank goodness!

A.2. Recruiting and compensating MTurk *workers*

Because Amazon MTurk is a labor marketplace, *worker* compensation can fluctuate widely. Initially, I paid \$1.00 for a 10-minute survey, which yielded few participants, and selected for participants who were on MTurk for a specific reason. \$0.20 and \$0.25 for a 20-minute survey (which was the price suggested by Paolacci et al. (2010)), yielded zero participants - the compensation was too low. Eventually, I settled on \$0.50 for ten minutes of “work.” The current price point is estimated to be about \$6.00 per hour (Paolacci et al., 2010).

Additionally, requiring Master qualifications raised the price significantly (both in terms of required compensation, and how much money Amazon MTurk demanded, while slowing the data collection. My recommendation for future experiments would be to restrict *worker* eligibility to 95% accuracy on hits, with a minimum of 1000 hits completed. Participants were also restricted to ones with an IP address within the United States, to try to control for native speakers of English. Hypothetically, you could set up a screening survey to test for native fluency, but that would likely drastically lower the number of participants willing to complete the actual experimental study. Because participants have generally been amiable, and keeping a high HIT accuracy rate is important, I was not too worried about dishonesty.

A.3. Notes on Using Qualtrics with MTurk

If you want to run a study with multiple blocks (or lists), there are several different ways to run a study on MTurk, using Qualtrics (an external survey platform). Because Qualtrics is a separate site, you must link your MTurk HITs to Qualtrics. Moreover, if you run multiple surveys on Qualtrics, you do not want the same *workers* taking multiple surveys.

A.3.1. Several surveys, one block at a time

You can set up one HIT for each block (each version of the experiment), and run all of your blocks simultaneously, running hundreds of participants overnight. In the instructions for each HIT, specify that “if you have already taken a survey similar to this for linguistics research, please do not complete this one.” In general, MTurk *workers* are genial and helpful; they realize that if they are caught ignoring the instructions, their HITs will not be approved.

At the end of the Qualtrics survey, the survey can generate a code (a random string of numbers) for the *worker* to input into the HIT as proof of completion (in order to be compensated; all HITs must be approved by the experimenter).³² In order to screen for *workers*, you must manually check for duplicates after surveys have been completed.

³²The most helpful set of instructions regarding Qualtrics and MTurk were found here: <http://reactiontimes.wordpress.com/2011/03/22/how-to-connect-qualtrics-and-mturk/>

This is not an ideal method, because you can only check for duplicates *after* the survey has been completed. However, this is a fairly quick method.

A.3.2. Screening Survey

Alternatively, Pe'er et al. (2012) suggest setting up one HIT with a screening survey (on Qualtrics). If an Amazon MTurk *worker* entered their Worker ID into the screening survey, and their Worker ID was already in your database of Worker IDs, they would not be able to take your Qualtrics survey. If their Worker ID was not in your database, the screening survey would redirect the *worker* to your actual survey. For example, your screening survey could link to Block A. Once you had run enough participants through Block A, you could set up your screening survey to link to Block B, and so on and so forth.

However, I found that there are a number of problems with this method. For one, you must run your experiment serially, one block at a time, which can take up more time than is ideal. For another, the process of redirection from the Amazon MTurk HIT interface to the Qualtrics screening survey, and then to your actual Qualtrics survey, is tremendously slow, and *workers* refuse to wait that long, especially when they are being compensated with so little.

A.3.3. One survey, in randomized blocks

In Qualtrics, though, it is possible to create one survey with multiple blocks, and to randomize within the survey. Additionally, it is possible to randomize equally across blocks (as opposed to having twice as many participants for Block A as you do for Block B, etc.). This one survey *does* become a little unwieldy, due to its sheer size, but this is the most efficient way to run multiple blocks simultaneously, while also ensuring that there are no duplicate *workers*, since you will have a one-to-one HIT-Qualtrics survey ratio. Mistakes are likely to be more costly, because the entire experiment will have to be rerun, as opposed to rerunning a block or two, but the advantages definitely outweigh the disadvantages with this method. It is recommended that you be very familiar with both the MTurk and Qualtrics interfaces before running a study in this manner.

B. Appendix: Comments by Participants

For the sake of amusement:

- “I could have sworn I saw a couple questions that went something like: ”Geographers do not realize that cats are insects.” For which there was no example answer provided. Presumably that is what you are studying?”
- “NO THOUGHTS”
- “Fun survey!”

- “This was very confusing to me and I’m a native speaker of American English but I had to read it over and over and I kept changing my answers. I’m still not sure but I did my best to comprehend what you were trying to ask. It’s a bit tricky to me. I enjoyed it though. Thanks!”
- “My kitty says hi.”
- “Semantics are a killer.”
- “This was tricky but great for a mental workout.”

C. Appendix: Study Forms

C.1. Informed Consent

My name is Paulina Haduong, and I am a senior Linguistics major at Yale University (BK '13). Please feel free to e-mail me at paulina (dot) haduong (at) yale (dot) edu if you have any comments, questions, or concerns. Your assistance with this research project is greatly appreciated.

All you need to do is complete this short questionnaire, which should take approximately 10 minutes. Responses will be completely anonymous; your name will not appear anywhere on the survey. Pressing "START" at the beginning of the survey constitutes your consent to participate. If you have any questions regarding your rights as a research participant, please contact the Institutional Review Board office at Yale University. Thank you for your help.

When you are ready to begin, please click "next" to proceed.

C.2. Cover Story

Aliens have invaded Earth!

In preparing for this invasion, different groups of aliens were trained as specialists in different things. One group was told everything there is to know about Earth geography. They are the Alien Geographers. A different group was told everything there is to know about Earth zoology (the animal life). They are the Alien Zoologists. Because they are aliens, they don't know anything about Earth unless they have specialized in it (Alien Zoologists have never heard of Africa, Alien Geographers have never of giraffes, etc.).

C.3. Stimuli

Item	Condition	Sentence
A01	a	Geographers do not realize that Paris is in Africa.
B01	b	Geographers do not realize that Paris is in Europe.
C01	c	Zoologists do not realize that Paris is in Europe.
D01	d	Geographers were told that Paris is in Europe.
E01	e	Geographers were told that Paris is in Africa.
A02	a	Geographers do not realize that London is in America.
B02	b	Geographers do not realize that London is in Europe.
C02	c	Zoologists do not realize that London is in Europe.
D02	d	Geographers were told that London is in Europe.
E02	e	Geographers were told that London is in America.
A03	a	Geographers do not realize that Rome is in Asia.
B03	b	Geographers do not realize that Rome is in Europe.
C03	c	Zoologists do not realize that Rome is in Europe.
D03	d	Geographers were told that Rome is in Europe.
E03	e	Geographers were told that Rome is in Asia.
A04	a	Geographers do not realize that Washington is in Africa.
B04	b	Geographers do not realize that Washington is in America.
C04	c	Zoologists do not realize that Washington is in America.
D04	d	Geographers were told that Washington is in America.
E04	e	Geographers were told that Washington is in Africa.
A05	a	Geographers do not realize that Boston is in Europe.
B05	b	Geographers do not realize that Boston is in America.
C05	c	Zoologists do not realize that Boston is in America.
D05	d	Geographers were told that Boston is in America.
E05	e	Geographers were told that Boston is in Europe.
A06	a	Geographers do not realize that Florida is in Asia.
B06	b	Geographers do not realize that Florida is in America.
C06	c	Zoologists do not realize that Florida is in America.
D06	d	Geographers were told that Florida is in America.
E06	e	Geographers were told that Florida is in Asia.
A07	a	Geographers do not realize that Beijing is in America.
B07	b	Geographers do not realize that Beijing is in Asia.
C07	c	Zoologists do not realize that Beijing is in Asia.
D07	d	Geographers were told that Beijing is in Asia.
E07	e	Geographers were told that Beijing is in America.
A08	a	Geographers do not realize that Delhi is in Africa.
B08	b	Geographers do not realize that Delhi is in Asia.
C08	c	Zoologists do not realize that Delhi is in Asia.
D08	d	Geographers were told that Delhi is in Asia.
E08	e	Geographers were told that Delhi is in Africa.

Item	Condition	Sentence
A09	a	Geographers do not realize that Tokyo is in Europe.
B09	b	Geographers do not realize that Tokyo is in Asia.
C09	c	Zoologists do not realize that Tokyo is in Asia.
D09	d	Geographers were told that Tokyo is in Asia.
E09	e	Geographers were told that Tokyo is in Europe.
A10	a	Geographers do not realize that Zimbabwe is in America.
B10	b	Geographers do not realize that Zimbabwe is in Africa.
C10	c	Zoologists do not realize that Zimbabwe is in Africa.
D10	d	Geographers were told that Zimbabwe is in Africa.
E10	e	Geographers were told that Zimbabwe is in America.
A11	a	Geographers do not realize that Ethiopia is in Europe.
B11	b	Geographers do not realize that Ethiopia is in Africa.
C11	c	Zoologists do not realize that Ethiopia is in Africa.
D11	d	Geographers were told that Ethiopia is in Africa.
E11	e	Geographers were told that Ethiopia is in Europe.
A12	a	Geographers do not realize that Cairo is in Asia.
B12	b	Geographers do not realize that Cairo is in Africa.
C12	c	Zoologists do not realize that Cairo is in Africa.
D12	d	Geographers were told that Cairo is in Africa.
E12	e	Geographers were told that Cairo is in Asia.
A13	a	Geographers do not realize that Nigeria is in America.
B13	b	Geographers do not realize that Nigeria is in Africa.
C13	c	Zoologists do not realize that Nigeria is in Africa.
D13	d	Geographers were told that Nigeria is in Africa.
E13	e	Geographers were told that Nigeria is in America.
A14	a	Geographers do not realize that Egypt is in Europe.
B14	b	Geographers do not realize that Egypt is in Africa.
C14	c	Zoologists do not realize that Egypt is in Africa.
D14	d	Geographers were told that Egypt is in Africa.
E14	e	Geographers were told that Egypt is in Europe.
A15	a	Geographers do not realize that Botswana is in Asia.
B15	b	Geographers do not realize that Botswana is in Africa.
C15	c	Zoologists do not realize that Botswana is in Africa.
D15	d	Geographers were told that Botswana is in Africa.
E15	e	Geographers were told that Botswana is in Asia.
A16	a	Zoologists do not realize that eagles are reptiles.
B16	b	Zoologists do not realize that eagles are birds.
C16	c	Geographers do not realize that eagles are birds.
D16	d	Zoologists were told that eagles are birds.
E16	e	Zoologists were told that eagles are reptiles.
A17	a	Zoologists do not realize that doves are mammals.
B17	b	Zoologists do not realize that doves are birds.
C17	c	Geographers do not realize that doves are birds.
D17	d	Zoologists were told that doves are birds.
E17	e	Zoologists were told that doves are mammals.

Item	Condition	Sentence
A18	a	Zoologists do not realize that seagulls are dogs.
B18	b	Zoologists do not realize that seagulls are birds.
C18	c	Geographers do not realize that seagulls are birds.
D18	d	Zoologists were told that seagulls are birds.
E18	e	Zoologists were told that seagulls are dogs.
A19	a	Zoologists do not realize that poodles are fish.
B19	b	Zoologists do not realize that poodles are dogs.
C19	c	Geographers do not realize that poodles are dogs.
D19	d	Zoologists were told that poodles are dogs.
E19	e	Zoologists were told that poodles are fish.
A20	a	Zoologists do not realize that dalmatians are birds.
B20	b	Zoologists do not realize that dalmatians are dogs.
C20	c	Geographers do not realize that dalmatians are dogs.
D20	d	Zoologists were told that dalmatians are dogs.
E20	e	Zoologists were told that dalmatians are birds.
A21	a	Zoologists do not realize that rottweilers are birds.
B21	b	Zoologists do not realize that rottweilers are dogs.
C21	c	Geographers do not realize that rottweilers are dogs.
D21	d	Zoologists were told that rottweilers are dogs.
E21	e	Zoologists were told that rottweilers are birds.
A22	a	Zoologists do not realize that trout are reptiles.
B22	b	Zoologists do not realize that trout are fish.
C22	c	Geographers do not realize that trout are fish.
D22	d	Zoologists were told that trout are fish.
E22	e	Zoologists were told that trout are reptiles.
A23	a	Zoologists do not realize that piranhas are insects.
B23	b	Zoologists do not realize that piranhas are fish.
C23	c	Geographers do not realize that piranhas are fish.
D23	d	Zoologists were told that piranhas are fish.
E23	e	Zoologists were told that piranhas are insects.
A24	a	Zoologists do not realize that swordfish are dogs.
B24	b	Zoologists do not realize that swordfish are fish.
C24	c	Geographers do not realize that swordfish are fish.
D24	d	Zoologists were told that swordfish are fish.
E24	e	Zoologists were told that swordfish are dogs.
A25	a	Zoologists do not realize that grasshoppers are mammals.
B25	b	Zoologists do not realize that grasshoppers are insects.
C25	c	Geographers do not realize that grasshoppers are insects.
D25	d	Zoologists were told that grasshoppers are insects.
E25	e	Zoologists were told that grasshoppers are mammals.
A26	a	Zoologists do not realize that beetles are fish.
B26	b	Zoologists do not realize that beetles are insects.
C26	c	Geographers do not realize that beetles are insects.
D26	d	Zoologists were told that beetles are insects.
E26	e	Zoologists were told that beetles are fish.

Item	Condition	Sentence
A27	a	Zoologists do not realize that caterpillars are dogs.
B27	b	Zoologists do not realize that caterpillars are insects.
C27	c	Geographers do not realize that caterpillars are insects.
D27	d	Zoologists were told that caterpillars are insects.
E27	e	Zoologists were told that caterpillars are dogs.
A28	a	Zoologists do not realize that cows are reptiles.
B28	b	Zoologists do not realize that cows are mammals.
C28	c	Geographers do not realize that cows are mammals.
D28	d	Zoologists were told that cows are mammals.
E28	e	Zoologists were told that cows are reptiles.
A29	a	Zoologists do not realize that bears are birds.
B29	b	Zoologists do not realize that bears are mammals.
C29	c	Geographers do not realize that bears are mammals.
D29	d	Zoologists were told that bears are mammals.
E29	e	Zoologists were told that bears are birds.
A30	a	Zoologists do not realize that giraffes are birds.
B30	b	Zoologists do not realize that giraffes are mammals.
C30	c	Geographers do not realize that giraffes are mammals.
D30	d	Zoologists were told that giraffes are mammals.
E30	e	Zoologists were told that giraffes are birds.
T01	b	Geographers do not realize that Italy is in Europe.
T02	c	Zoologists do not realize that Germany is in Europe.
T03	d	Geographers were told that Ireland is in Europe.
T04	e	Geographers were told that Las Vegas is in Africa.
T05	e	Geographers were told that New Orleans is in Asia.
T06	d	Geographers were told that New York is in America.
T07	c	Zoologists do not realize that Hong Kong is in Asia.
T08	b	Geographers do not realize that China is in Asia.
T09	e	Geographers were told that Malaysia is in America.
T10	b	Geographers do not realize that Johannesburg is in Africa.
T11	c	Zoologists do not realize that Kentucky is in America.
T12	d	Geographers were told that the Sahara is in Africa.
T13	e	Zoologists were told that chickens are fish.
T14	b	Zoologists do not realize that parrots are birds.
T15	c	Geographers do not realize that chihuahuas are dogs.
T16	d	Zoologists were told that bulldogs are dogs.
T17	e	Zoologists were told that catfish are insects.
T18	c	Geographers do not realize that goldfish are fish.
T19	d	Zoologists were told that ants are insects.
T20	e	Zoologists were told that ants are birds.
T21	b	Zoologists do not realize that foxes are mammals.
T22	c	Geographers do not realize that rabbits are mammals.
T23	d	Zoologists were told that rattlesnakes are reptiles.
T24	b	Zoologists do not realize that geckos are reptiles.

References

- Abrusán, Márta, and Kriszta Szendrői. 2012. Experimenting with the king of France. In *Logic, language and meaning*, ed. Maria Aloni, Vadim Kimmelman, Floris Roelofsen, Galit W. Sassoon, Katrin Schulz, and Matthijs Westera, number 7218 in Lecture Notes in Computer Science, 102–111. Springer Berlin Heidelberg.
- Atkinson, George W. 1998. *Chess and machine intuition*. Intellect Books.
- Atlas, Jay. 2005. *Logic, meaning, and conversation: Semantical underdeterminacy, implicature, and their interface*. {Oxford University Press}.
- Behrend, Tara S, David J Sharek, Adam W Meade, and Eric N Wiebe. 2011. The viability of crowdsourcing for survey research. *Behavior research methods* 43:800–813. PMID: 21437749.
- Berinsky, Adam J., Gregory A. Huber, and Gabriel S. Lenz. 2011. Using Mechanical Turk as a subject recruitment tool for experimental research. *Submitted for review* .
- Buhrmester, Michael, Tracy Kwang, and Samuel D. Gosling. 2011. Amazon’s Mechanical Turk a new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science* 6:3–5.
- Chemla, Emmanuel, and Lewis Bott. 2010. Processing presuppositions: Dynamic semantics vs pragmatic enrichment. *Language and Cognitive Processes* 0:1–20.
- Donnellan, Keith S. 1966. Reference and definite descriptions. *The philosophical review* 75:281–304.
- Fodor, Jerry D. 1979. In defense of the ‘truth value gap’. *Syntax & Semantics* 11:199–224.
- Goodman, Joseph K., Cynthia E. Cryder, and Amar Cheema. 2012. Data collection in a flat world: The strengths and weaknesses of Mechanical Turk samples. *Journal of Behavioral Decision Making* .
- Gosling, Samuel D, Simine Vazire, Sanjay Srivastava, and Oliver P John. 2004. Should we trust web-based studies? a comparative analysis of six preconceptions about internet questionnaires. *The American psychologist* 59:93–104. PMID: 14992636.
- Grice, H. Paul. 1975. Logic and conversation. *1975* 41–58.
- Heim, Irene. 1983. On the projection problem for presuppositions. *Formal Semantics: The Essential Readings* .
- Horn, Laurence R. 1991. Given as new: When redundant affirmation isn’t. *Journal of Pragmatics* 15:313–336.
- Ipeirotis, Panagiotis. 2010. Demographics of Mechanical Turk. SSRN Scholarly Paper ID 1585030, Social Science Research Network, Rochester, NY.
- Karttunen, Lauri. 1973. Presuppositions of compound sentences. *Linguistic Inquiry* 4:169–193.
- Karttunen, Lauri. 1974. Presupposition and linguistic context. *Theoretical Linguistics* 1:181–194.

- Kiparsky, Paul, and Carol Kiparsky. 1970. *Fact*. Linguistics Club, Indiana University.
- Lewis, David. 1979. Scorekeeping in a language game. *Journal of Philosophical Logic* 8:339–359.
- Paolacci, Gabriele, Jesse Chandler, and Panagiotis Ipeirotis. 2010. Running experiments on Amazon Mechanical Turk. *Judgment and Decision Making* 5:411–419.
- Parsons, Terence. 1990. True contradictions. *Canadian Journal of Philosophy* 20:335–353. ArticleType: research-article / Full publication date: Sep., 1990 / Copyright © 1990 Canadian Journal of Philosophy.
- Pe'er, Eyal, Gabriele Paolacci, Jesse Chandler, and Pam Mueller. 2012. Selectively recruiting participants from Amazon Mechanical Turk using Qualtrics. SSRN Scholarly Paper ID 2100631, Social Science Research Network, Rochester, NY.
- Priest, Graham. 1987. *In contradiction*. Dordrecht: M. Nijhoff.
- Russell, Bertrand. 1905. On denoting. *Mind* 14:479–493.
- Schlenker, Philippe. 2008. Be articulate: A pragmatic theory of presupposition projection. *Theoretical Linguistics* 34:157–212.
- Simons, Mandy. 2007. Presupposition and cooperation. *Unpublished Manuscript* .
- Smith, E. Allyn, Mayol, Laia, and Castroviejo, Elena. 2013. Felicity of direct denial by meaning type in english.
- Smith, Nicholas J. J. 2008. *Vagueness and degrees of truth*. Oxford University Press.
- Stalnaker, Robert. 1999. *Context and content: Essays on intentionality in speech and thought*. Oxford University Press.
- Stigen, Anfinn. 1960. What does Mr. Tennessen mean, and what should i say? *Inquiry* 3:180–184.
- Strawson, P. F. 1964. Identifying reference and truth-values. *Theoria* 30:96–118.
- Strawson, Peter Frederick. 1950. On referring. *Mind* 59:320–344.
- Tennessen, Herman. 1960. Vindication of the Humpty Dumpty attitude towards language. *Inquiry* 3:185–198.
- Thomason, Richmond. 1990. Accommodation, meaning, and implicature: Interdisciplinary foundations for pragmatics. *Intentions in communication* 325363.
- Von Stechow, Kai. 2004. Would you believe it? the king of France is back! presuppositions and truth-value intuitions. *Descriptions and beyond* 315–341.
- Von Stechow, Kai. 2008. What is presupposition accommodation, again?*. *Philosophical Perspectives* 22:137–170.
- Ward, Gregory. 2003. Comments on “Felicity and Presupposition Triggers” by Michael Glanzberg. University of Michigan.
- Yablo, Stephen. 2006. Non-catastrophic presupposition failure. *Content and Modality: Themes from the Philosophy of Robert Stalnaker*, Clarendon Press, Oxford 164–190.

Yalcin, Seth, and Joshua Knobe. 2010. Fat Tony might be dead: An experimental note on epistemic modals. *online at: <http://www.semanticsarchive.net/Archive/TdiZjA3N>*