# Introduction

## Referential uncertainty

A challenge for word learners, known as “referential uncertainty”, is to map unknown words onto their referents (Quine, 1960). To illustrate the problem, imagine a young child who has in front of her a bowl of peas, a bowl of carrots, and a bowl of mashed potatoes. She does not know the names for any of these foods. If her parent says, “Eat your peas,” the child does not know whether the word “peas” refers to the green food, the orange food, or the white food. As Quine points out, the word “peas” might even refer to part of a food, like a single pea or a single carrot; or to vegetables or food in general; or to the bowl or the table; or even to something in another linguistic category, such as the adjective *orange* or the action of eating. How do word learners, including both children and adults, map a word to its referent object given the problem of referential uncertainty?

The following introductory section traces some of the responses in the literature to the problem of referential uncertainty, focusing on the emergence and development of cross-situational statistics as a potential way to solve this word learning problem. I first briefly mention some early work that proposed ways in which referential uncertainty might be resolved in single naming events. I then discuss the more recent literature on cross-situational word learning and, in particular, the debate about the specific mechanisms that might underlie a cross-situational solution to referential uncertainty. I consider recent accounts, including this study’s immediate predecessors, that integrate early social theories of word learning with cross-situational theories to shed light on this debate. Finally, I examine the ways in which eye-tracking has been used in word-learning studies. In the present study, it is used in a novel way to further illuminate the underlying mechanisms of cross-situational word learning.

**Subsection 1**

Many researchers studying word learning between the late 1980s and the early 2000s focused on single naming events. As the term suggests, a single naming event is a singular instance in which a learner hears a novel word and must at that time identify the word’s referent, a process known as “fast mapping” (Yu & Smith, 2007). Though there could be an infinite amount of possible referents for a given novel word, some researchers have suggested the existence of constraints that might help to narrow the possibilities and help the learner map the word to its referent. Markman (1990) posits that even very early word learners might hold conceptual assumptions that bias them towards certain potential referents of novel words. For instance, she argues that the “whole object assumption” leads children to find it more likely that a novel word refers to an entire object instead of to part of that object or to one of its qualities. Gleitman (1990) suggests a process called “syntactic bootstrapping”, through which the syntactic structure of an utterance places constraints upon the space of possible referents. In the above example, a word learner who is familiar with the meaning or use of the word “eat” can infer from the beginning of the command that “peas” are a type of food; this would rule out, for instance, the hypothesis that “peas” referred to the bowl, since bowls are inedible.

**[SOCIAL INFO]**

**-Baldwin, Bloom, Kachergis**

## Cross-situational statistical word learning

In contrast to single naming events, cross-situational statistical word learning involves the aggregation of information across multiple word learning events or *situations*. Learners might use statistical information about the frequency or the distribution of words and objects to infer word-referent pairs over time. Consider the above example: if the next day the child’s parent were to say, “Eat your peas,” over a dinner of peas, rice, and banana, the child would have more information to indicate a link between “peas” and the green food, since the word “peas” has co-occurred with that object more often than with the others.

There are a number of reasons to suspect that word learners, and young children in particular, make use of cross-situational word learning. Even taking into account the constraints discussed above, single naming events in real life are often noisy and ambiguous (Yu & Smith, 2007; Smith & Yu, 2008). In addition to there often being many possible referents for a given novel spoken word, one must also take into account referents in a number of categories (not just nouns; also adjectives, verbs, etc.); multi-word referents; and referents that are either not physically present or completely intangible (like *love*) (Medina, Snedeker, Trueswell, & Gleitman, 2011). Additionally, while adults can make recourse to syntactic, lexical, pragmatic, and other contextual clues when learning the meaning of a new word (for example, using Gleitman’s proposed process of *syntactic bootstrapping*), young infants have not all acquired the information needed to rely on such clues (Siskind, 1996; Smith & Yu, 2008; Smith, Suanda, & Yu, 2014). Cross-situational learning might be an answer to this problem. Finally, infants have been shown to use statistical information to differentiate spoken words within spoken sentences (Saffran et al., 1996), so it is plausible that a similar statistical mechanism might be used to map novel words to referents.

Both adults and children have demonstrated the ability to learn word-object pairs across multiple ambiguous naming situations (Yu & Smith, 2007; Smith & Yu, 2008; Voulousmanos, 2008; Kachergis, Yu, & Shiffrin, 2014; Escudero, Mulak, & Vlach 2015; Yu & Smith 2011). Prior to this time, formal computation models such as Siskind’s (1996) had suggested that cross-situational statistics were a viable solution to the problem of referential uncertainty, but Smith and Yu’s 2007 adult study was the first to examine cross-situational word learning in human learners. Because subsequent studies, including the present one, have presented variations on the basic design used in Smith and Yu’s 2007 and 2008 cross-situational word learning experiments, I briefly outline the paradigm here. Participants saw training trials consisting of several pictures of objects and simultaneously heard several novel words corresponding to the number of objects. The words and objects were repeated over subsequent trials, with each word presented always in conjunction with the same object. No other information was given that might indicate the object to which a word mapped. Test trials at the end of the experiments measured whether participants had learned which words referred to which objects via forced-choice tests for adults (2007) and manually-coded looking times for children (2008). Both age groups learned more word-object relationships than would have been expected by chance.

There remain important questions about the potential use of cross-situational word learning. Recently, Smith, Suanda, and Yu (2014) questioned whether infants are able to learn words cross-situationally outside of laboratory settings, in noisier contexts. [add more here]

Yu and Smith (2007, 2008) propose two ideas for the mechanisms that learners use to arrive at the conclusion that a word refers to a certain object across situations. The first idea, here referred to as *multiple-alternative tracking*, is that learners track multiple alternatives for the possible referent, and the second, *single-hypothesis tracking*, is that learners store a single strong hypothesis about the object a word refers to. Much of the literature since then, including the present study, has focused on distinguishing between these processes, with recent attention directed toward the ways in which they might interact. In the immediately following section, I will explain some of the arguments and evidence that have been put forth first for multiple-alternative tracking, then for single-hypothesis tracking, and finally for integrative accounts of both ideas.

## Competing accounts of cross-situational word learning processes

### Multiple-alternative tracking

Multiple-alternative tracking, also known as “associative learning” (Smith, Suanda, & Yu, 2014), posits that learners keep track of multiple possible referents for a given word across naming situations. Specifically, the learner might track information about the statistical distribution of the possible referents, including how often the object occurs in the same setting as the spoken word. The account predicts that after some number of naming situations, one object will emerge as having co-occurred with the word more often than other objects; when asked about a word’s referent, learners will then choose this object. To illustrate, consider again our example of the child and the peas. If the child is tracking multiple alternatives, she remembers all of the possible things “peas” might refer to. For instance, if the first time she hears “peas” she sees a green, an orange, and a white food, and the second time she sees a green, an orange, and a yellow food, then she might expect the green or the orange food to be twice as likely to be the referent of “peas” than is the white or the yellow food. Eventually, after a number of such naming instances, the green food will appear to the child statistically most likely to be the referent of “peas”.

The multiple-alternative account predicts that word learners will demonstrate memory of multiple potential referents of a given word, even those that are not the most likely referent. Consistent with this prediction, Vouloumanos (2008) found that adult word learners differentiated between objects that were, for instance, 20% likely to be the referent of a word versus objects that were 10% likely to be the referent, even if neither object was statistically the *most* likely. This finding suggests that learners retain information about objects beyond simply the most likely referent. Kachergis, Yu, and Shiffrin (2014) found a similar phenomenon, noting that adult learners kept track of co-occurrences between the word and multiple objects, although this result was stronger when participants were explicitly asked to learn the meanings of words compared to an implicit task that did not contain that instruction.

[more associative work: Yu 2008; McMurray, Samuelson, & Horst 2012; Yurovsky, Boyer, Smith, & Yu 2012]

### Single-hypothesis tracking

Proponents of single-hypothesis tracking, also known as “propose-but-verify” (Woodard, Gleitman, & Trueswell, 2016; Trueswell et al., 2013) or “hypothesis testing” (Smith, Suanda, & Yu, 2014), argue that learners form a single hypothesis about a word’s referent on the first naming events. Over subsequent naming events – that is, cross-situationally – the hypothesis is either strengthened by the continued co-occurrence of the word and hypothesized referent, or rejected and replaced with a new hypothesis. In the peas example, if the child is tracking a single hypothesis, she might hypothesize the first time she hears the word “peas” that it refers to the orange food. If the orange food is present when she next hears “peas”, her hypothesis is strengthened. If it is absent, she must form a new hypothesis.

Single-hypothesis tracking bears a strong resemblance to the “fast mapping” approach discussed above, which research suggests is used in single naming events. To clarify, the single-hypothesis tracking account *combines* fast mapping on the first naming event (the “proposal” of the hypothesis) with amendments to the mapping on subsequent naming events (the “verification” of the hypothesis). As mentioned earlier in this introduction, a number of studies have shown that word learners use various conceptual, pragmatic, and linguistic constraints to correctly map some words to referents on single naming events (e.g. Carey, 1978; Baldwin, 1993). The observance of this ability provides some evidence for single-hypothesis tracking accounts of cross-situational word learning.

Other arguments for the use of single-hypothesis tracking in cross-situational word learning come from Medina et al. (2011). In an attempt to more faithfully represent real-world word learning situations than in previous word-learning studies, they asked adults to watch muted vignettes of parents speaking to children, with one word in the vignette replaced by a novel nonsense word. Participants were asked what they thought the meaning of each nonsense word was after each of five vignettes, to track whether their hypotheses changed over time. Medina et al. found, contrary to what might be predicted by associative accounts, that participants’ accuracy in mapping words to referents did not improve across trials. They also observed that accuracy on the final guess depended on how informative the *first* trial had been, a finding consistent with the fast mapping process supposedly involved in single-hypothesis tracking.

Both adults and children have been shown to engage in behaviors expected of a single-hypothesis tracker (Medina et al., 2011; Trueswell et al., 2013; Woodard, Gleitman, & Trueswell, 2016). Trueswell et al. (2013) and Woodard, Gleitman, and Trueswell (2016) make use of an additional metric that is also used in the current study to identify behavior consistent with single-hypothesis tracking. [these studies use a setup that’s similar to the same/switch one in MacDonald & in the present study – not sure if I should explain them here?] A word learner who forms a single strong hypothesis might be predicted to fail to remember the other possible referents. Indeed, these studies found that both adults and children performed at chance when asked to identify the referent of a word from a set of objects that did not include their (incorrectly) hypothesized referent.

### Integrative accounts

While earlier research advocated for one or the other underlying mechanism of cross-situational word learning, recent work has focused on integrative accounts of the two. Yurovsky and Frank (2015) found not only that learners appear to use both mechanisms during cross-situational word learning, but also that certain variables can be manipulated to make learners look more like either single-hypothesis or multiple-alternative trackers. In particular, the study manipulated both the number of potential referents for a given word (either 2, 3, 4, or 8) and the intervening trials between the learning and the test trial for a word (either 1, 2, 4, or 8). When more potential referents were presented, learners were more likely to behave like single-hypothesis trackers and had worse memory for the objects they had not hypothesized as a word’s referent. In contrast, when fewer objects were presented, learners were better able to track multiple alternative potential referents. This finding suggests that the two ideas fall on a continuum and that learners employ different strategies based on their uncertainty during the learning situation.

MacDonald, Yurovsky, and Frank (2015) altered the paradigm used in Yurovsky and Frank (2015) to include the presence of a social cue, namely the speaker’s eye gaze. As noted above, social information has been shown to modulate uncertainty in single naming events (Baldwin, 1991; Baldwin & Moses, 2001). MacDonald, Yurovsky, and Frank found that when adult participants watched a face turn its head toward one of the potential referent objects while hearing a nonsense word spoken, they were less likely to track multiple alternative possible referents and more likely to behave as if they had stored a single hypothesis.

## Eye-tracking

The present study aims to replicate the key finding from MacDonald, Yurovsky, and Frank (2015) – namely, that the presence of eye gaze causes adult word learners to behave more like single-hypothesis trackers and less like multiple-alternative trackers. Unlike this study, however, the present study uses an eye-tracker both to examine participants’ behavior while learning words and to identify their confidence in the referent of a word during a testing phase. Eye-tracking data is very precise (Huettig & Altmann, 2005), allowing for timecourse analyses of participant gaze. Further, using eye gaze as the indicator of the participant’s hypothesized referent removes the need for participants to make a forced choice, which obscures the participant’s potential uncertainty about the object she thinks the word refers to. Finally, eye-tracking allows for the study of younger populations for whom the word-learning question is more relevant (Fernald, Zangl, Portillo, & Marchman, 2008; Halberda, 2006). Eye-tracking has recently been used reliably to study cross-situational word learning in young children (Yu & Smith, 2011), although it has not yet been used to study the presence of social cue in cross-situational word learning, nor has it been used with the paradigm described in the section on integrative accounts above. While the present study examines only adult word learning, I hope it will inform future work that seeks to use eye-tracking data to analyze the interaction between social cue and cross-situational word learning in children.

[revisit this section & flesh out]

## Hypotheses

I predict, based on findings from MacDonald, Yurovsky, and Frank (2015), that the presence of a speaker’s eye gaze will reduce referential uncertainty and cause participants to behave more like single-hypothesis trackers. Empirically, I predict to see that participants who see the social cue on learning trials will allocate more attention to the target of the speaker’s gaze than to other potential referents, and that participants who do not see the social cue will allocate attention equally across potential referent objects. I also predict to see a quantitative relation between a learner’s success at identifying a word’s referent during the testing phase and the time spent looking at that object during the learning phase. Finally, I predict that participants who see the social cue during the learning phase will perform differently during the testing phase than those who did not see a social cue, even if both participants allocated the same amount of attention to an object during the learning phase. [this is probably confusing – reword]