

Talking in Another Person's Shoes: Incremental Perspective-taking in Language Processing

Sarah Brown-Schmidt

BROWNSCH@ILLINOIS.EDU

*Department of Psychology and Beckman Institute
University of Illinois at Urbana-Champaign
Champaign, IL, 61820, USA*

Joy E. Hanna

JOY.HANNA@OBERLIN.EDU

*Department of Psychology
Oberlin College
Oberlin, OH, 44074, USA*

Editors: David Schlangen, Hannes Rieser

Abstract

Language use in conversation is fundamentally incremental, and is guided by the representations that interlocutors maintain of each other's knowledge and beliefs. While there is a consensus that interlocutors represent the perspective of others, three candidate models, a Perspective-Adjustment model, an Anticipation-Integration model, and a Constraint-Based model, make conflicting predictions about the role of perspective information during on-line language processing. Here we review psycholinguistic evidence for incrementality in language processing, and the recent methodological advance that has fostered its investigation—the use of eye-tracking in the visual world paradigm. We present visual world studies of perspective-taking, and evaluate each model's account of the data. We argue for a Constraint-Based view in which perspective is one of multiple probabilistic constraints that guide language processing decisions. Addressees combine knowledge of a speaker's perspective with rich information from the discourse context to arrive at an interpretation of what was said. Understanding how these sources of information combine to influence interpretation requires careful consideration of how perspective representations were established, and how they are relevant to the communicative context.

Keywords: conversation, eye-tracking, perspective, processing

1 Introduction

Recent years have seen a surge of interest in how language is processed in real time and in real space. This interest has, in part, been stimulated by the advent of modern eye-tracking devices that can monitor moment-by-moment language processing in settings that mimic many aspects of real world language use. These experimental techniques have enabled researchers to address central questions about the incrementality of the language processing system. Words occur over time, at an average of 150-200 words per minute (Tauroza & Allison, 1990; also see Levelt, 1989, p. 22). As a result, many words and phrases are ambiguous, at least temporarily, causing a proliferation of possible syntactic and semantic interpretations as they accrue. Thus a critical problem for theories of language processing is to explain how listeners build a representation of an on-going sentence when the meaning of many words is unclear until later in the sentence. Tanenhaus (2004, p. 376) describes (and ultimately criticizes) an early view of language understanding as a catch-up game in which the listener held each word in a memory buffer, waiting for subsequent information that could disambiguate the words. Instead, evidence from over a decade of studies using on-line methods have revealed that: (a) words are integrated

immediately into the on-going sentence, with listeners making provisional commitments to interpretations of lexical (Allopenna, Magnuson, & Tanenhaus, 1998; Dahan, Magnuson, Tanenhaus, & Hogan, 2001), referential (Hanna, Tanenhaus & Trueswell, 2003; Chambers et al., 2002), syntactic (Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995), and other ambiguities; and (b) listeners make sophisticated predictions about upcoming material (Altmann & Kamide, 1999, 2007, 2009; De Goede et al., 2009). This evidence has led to further work addressing questions about the timing with which various sources of linguistic and non-linguistic information, from lexical frequency and neighborhood effects (Magnuson, Dixon, Tanenhaus, & Aslin, 2007) to speaker identity (Brown-Schmidt, 2009a; Metzing & Brennan, 2003), are used during the incremental process of language use.

In this chapter, we focus on the use of one type of non-linguistic information in incremental language comprehension—information about others' knowledge and beliefs. As we shall see, studying perspective-taking in language processing requires methodologies that afford the use of naturally produced speech in rich, interactive contexts, while at the same time providing detailed information about the time-course of production and comprehension. In the last 10-15 years, the application of eye-tracking methodology to dialog tasks has stimulated significant debate and theorizing about when considerations of the speaker's perspective play a role during processing. Three primary candidate models of the role of perspective during on-line comprehension have emerged: a Perspective-Adjustment model (Keysar, Barr, Balin, & Paek, 1998), an Anticipation-Integration model (Barr, 2008), and a Constraint-Based model (Hanna, et al., 2003). As we shall see, the key differences between these models are in their proposals for when perspective guides interpretation decisions. In order to set the stage for our discussion of this issue, and our perspective on it, the remainder of this section will provide an overview of how words are interpreted incrementally in rich contexts. We then turn to the question of how knowledge of speaker perspective modulates this process.

Much of the support for incremental interpretation of language comes from studies that use the visual world paradigm (VWP) developed by Tanenhaus and his colleagues (Tanenhaus, et al., 1995; also see Cooper, 1974; Pechmann, 1989). In the typical visual-world experiment, a participant addressee is presented with a display (either in the real world or on a computer screen) and is asked to execute a command that involves manipulating objects in the display. Meanwhile, her eye movements are monitored, either with a head-mounted or a remote eye-tracking device, at a resolution between 30 and 2000 samples per second, depending on the type of equipment. People typically freely attend to and look at objects as they are mentioned (Tanenhaus, Magnuson, Dahan, & Chambers, 2000), particularly when they have to reach for or click on the objects. Further, listeners tend to gaze at objects that share semantic (Yee & Sedivy, 2006) or visual features (Dahan & Tanenhaus, 2005; Huettig & Altmann, 2005) with a mentioned object (Altmann & Kamide, 2007). Thus, eye movements provide a relatively unobtrusive measure of an addressee's linguistic interpretation. Researchers can calculate the timing and duration of looks to the various objects in a display, and measures include the amount of time it takes before a participant fixates on the target object for the first time, the amount of time it takes before a participant fixates on the target object for the last time (usually just before they select the object), and the changes over time in the likelihood of looking at the target and other potential target objects (known as competitors), as the critical linguistic expression unfolds. The VWP initially focused on the participant addressee's comprehension processes while following directions from a speaker, who very often was the experimenter or a confederate participant who was using a script or trained in what to say. Other work extended this paradigm to language production (Griffin & Bock, 2000; Meyer, Sleiderink, & Levelt, 1998), and more recently to the comprehension of spontaneous, interactive speech (Hanna & Brennan, 2007; Brown-Schmidt & Tanenhaus, 2008; Brown-Schmidt, Campana, & Tanenhaus, 2005).

The typical design in the VWP makes use of definite referring expressions (e.g., *the starred yellow square*), that refer to a single unique object in the display, but that are temporarily ambiguous among several objects when produced over time by the speaker (e.g., when there are several starred objects). These studies often manipulate and examine the point of disambiguation (POD) of the referring expression; that is, the point at which the interpretation of the expression is disambiguated given linguistic or other information available to the addressee in the visual world of the experiment. This methodology has not only allowed researchers to examine the basic nature of the incremental interpretation of noun and other linguistic phrases, but has also proven to be a powerful tool for examining the influence of higher-level sources of information on reference resolution, such as the pragmatics of how the objects in the display can be manipulated (Chambers et al., 2002; Hanna & Tanenhaus, 2004) as well as information from the context of being engaged in a conversation, such as the knowledge or perspective of the speaker and whether it matches that of the addressee (Keysar, Barr, Balin, & Brauner, 2000; Hanna, et al., 2003). We will further explore the latter work in this paper, but will first describe an example VWP experiment that illustrates the fundamentally incremental nature of language processing.

In one of the first studies to use the VWP, Eberhard, Spivey-Knowlton, Sedivy, and Tanenhaus (1995) showed participants displays with four objects, including colored squares and rectangles, some of which were starred. On critical trials, participants heard instructions to *Touch the starred yellow square*. The critical manipulation was when, during the noun phrase (underlined), the display afforded unique identification of the target. When the display contained only a single starred object, the target could be identified at the word *starred*. When all four objects were starred, but only one was yellow, the target could be identified at *yellow*. In a final condition, the display contained a starred yellow square (the target) and a starred yellow rectangle, along with two different color shapes, meaning that the target could only be identified at the final word *square*. Eberhard, et al.'s experiment allowed for a critical test of the incrementality hypothesis: if listeners interpret the test sentence incrementally, narrowing in on an identification of the target referent with each new word, then there should be significant differences in how quickly listeners look at the target referent in the three conditions. The results were consistent with this hypothesis; listeners were significantly faster to identify the target when the scene provided an earlier point of disambiguation. Reference resolution also happened incredibly quickly, in a fashion that was incrementally time-locked to the words in the noun phrase. Taking into account the fact that it takes about 200 ms to program and launch a saccadic eye movement (Hallet, 1986), looks to the target object were programmed immediately after or even before the end of the disambiguating word.

The Eberhard, et al. findings demonstrate that during language processing, interpretation takes place incrementally, taking into account both the unfolding linguistic input as well as the domain of objects against which that input is being interpreted. It also means that information that has the potential to change the domain of interpretation (or referential domain, cf. Salmon-Alt & Romary, 2000 or Chambers, et al., 2002) can be manipulated and examined within the VWP as well. This is important, because the process of understanding a sentence is more complex than simply matching words onto entities in the world. Instead, addressees use a variety of sources of information to resolve ambiguity and form meaningful representations of sentences as they unfold in time. Employing these sources of information to constrain language processing may require accessing information from long-term memory, holding multiple pieces of information in working memory, or complex inferencing procedures. Thus, a critical question is whether the language processing system is able to access and use various sources of information at a speed quick enough to affect interpretation processes as they occur in real time. The diagnostic of these effects is the same as it was for the first demonstrations of incrementality: determining whether a given source of information can affect the domain of interpretation for a referring expression by including or excluding candidate referents, and therefore changing the point of disambiguation

for that expression. This area of research has generated many noteworthy findings, among them the fact that all of the following factors constrain the domain of interpretation incrementally with each successive word (or partial word, see Dahan, et al., 2001): action affordances (Chambers, et al., 2002); verb semantic restrictions (Dahan & Tanenhaus, 2004; Altmann & Kamide, 1999); fluency of referential expressions (Arnold, Hudson Kam, & Tanenhaus, 2007); scalar implicatures (Sedivy, Tanenhaus, Chambers, & Carlson, 1999); speaker reliability (Grodner & Sedivy, in press); speaker identity (Metzing & Brennan, 2003); speaker eye gaze (Hanna & Brennan, 2007); and event knowledge (Nieuwland, Otten, & Van Berkum, 2007). Below we consider in detail one factor that has been proposed to fundamentally define the domain of interpretation for interlocutors in conversational settings – knowledge about what information is and is not in common ground.

2 Common Ground and Perspective-taking

Common ground is the set of mutual knowledge and beliefs shared among conversational participants, and is thought to be formed on the bases of community membership, linguistic interactions (known as linguistic co-presence), and physical environments (known as physical co-presence) (Clark, 1992; 1996; Clark & Marshall, 1978, 1981). A large body of work by Herbert Clark and his colleagues has focused on the nature of common ground, which they view as a core component of language use, affecting many basic aspects of the form and process of reference (e.g., Clark & Wilkes-Gibbs, 1986; Wilkes-Gibbs & Clark, 1992). Successful communication often requires keeping track of what is and is not in common ground with your conversational partner; for example, the experimenter in Eberhard et al. (1995) could not successfully refer to *the starred yellow square* if the experimental participant could not see that shape in the display. Indeed, Clark has argued that common ground defines the relevant domain of interpretation (Clark, 1992; 1996), and as such plays a central role in language processing. However, the question of whether common ground can restrict the domain of interpretation incrementally and rapidly, and the degree to which it restricts possible interpretations, was relatively unexplored until recently, and remains controversial.

This is partly due to the methods available for studying language use in conversation. The typical common ground experiment employs some type of referential communication task (based on the original by Krauss & Weinheimer, 1966), where two people must work together to construct or arrange objects in a display without being able to see each other or each other's display, although they are able to communicate freely in ways that maintain many aspects of naturally occurring conversation. Usually, the speaker or director knows the goal arrangement of objects, and gives directions to the addressee or matcher to move and manipulate them in order to get their displays to match. Before the advent of head-mounted eye tracking devices, researchers using referential communication tasks were limited to measures of language processing on a relatively coarse time-scale, such as the number of speaking turns taken by the director and matcher, the number of confusions or errors indicated by an incorrect object selection, and the total time it took to complete the task (e.g., see Clark & Wilkes-Gibbs, 1986). While these measures provided a wealth of data about the nature of common ground, they were not able to provide information about an addressee's interpretation on the moment-by-moment basis needed to assess the effects of common ground on reference resolution and domain restriction. In addition, prior to the development of the VWP, the techniques that were available for studying on-line processing, such as reading-time measures, or event-related potentials, were not easily adaptable for use with real-world contexts and in conversational settings.

The role of common ground in language processing was also neglected due to the preponderance of processing models that limited the hypothesized role of contextually based sources of information, in order to explain how a complex process such as language

comprehension could occur so rapidly and seemingly effortlessly. Early models of syntactic parsing proposed that this efficiency is due to the encapsulation of syntactic processes from other sources of information, such as discourse context, which were thought to require resource-intensive processing (Ferreira & Clifton, 1986; see discussion in Trueswell & Tanenhaus, 1994). Sentences were thought to be processed first by a fast, syntactic parser, with more complex sources of information, such as the number of referents in the discourse context, playing a role only during a later revision stage. This early view was in direct conflict with the view of Clark and colleagues that the context in which language occurred was central to language itself (Clark, 1992, 1996).

The crucial turning point in our understanding of the role of context in real-time language comprehension was the finding by Tanenhaus, et al., (1995), that rich information from a scene can eliminate or dramatically reduce (see Novick, Thompson-Schill, & Trueswell, 2008) syntactic ambiguity. This, and subsequent experiments using the VWP (Chambers, Tanenhaus, & Magnuson, 2004; Chambers, et al., 2002; Spivey, Tanenhaus, Eberhard, & Sedivy, 2002; Trueswell, Sekerina, Hill, & Logrip, 1999; Novick, et al., 2008; Eberhard, et al., 1995) have demonstrated that as listeners interpret an utterance, information from the discourse context, such as the number and features of the potential referents, immediately constrains the interpretation of the sentence, and that there is clearly no early, context-free processing stage. Subsequent research extended these findings, demonstrating that contextual information is combined with a variety of sources of information including verb bias, thematic fit, prosody, etc. to guide language processing decisions over time (e.g., Beun & Cremers, 1998; Brown-Schmidt & Tanenhaus, 2008; Niewland, et al., 2007; Hanna & Tanenhaus, 2004; Arnold & Griffin, 2007; Metzing & Brennan, 2003; Wilson & Garnsey, 2009, Watson, Tanenhaus, & Gunlogson, 2008, Arnold, et al., 2007; Dahan & Tanenhaus, 2004; Sedivy, et al., 1999; Altmann & Kamide, 1999).

If common ground is central to language (Clark, 1992, 1996; Clark & Wilkes-Gibbs, 1986), and serves as the primary context for language understanding, then the clear prediction would be that common ground guides the incremental processing of language. As we shall see, soon after the development of the VWP, researchers began to extend the paradigm to examine whether common ground constrained referential processing (e.g., Keysar, et al., 1998). Before we review this literature, it is important to first consider what the role of common ground would be. For every pair of individuals, some of their knowledge is likely to be shared, and some of their knowledge will be private. One role might be using information about what is shared and what is private to circumscribe the domain of interpretation of a referring expression. For example, if a speaker were to say *That's a lovely ring!*, the addressee should constrain the domain of interpretation to a ring that is visible to both speaker and addressee, and not, say, a toe-ring concealed by the addressee's shoe. Likewise, if the speaker were to ask *What just happened?*, a cooperative addressee (Grice, 1975) would be expected to provide information not already known to the speaker.

As we shall see, studies of the role of common ground in incremental language understanding typically employ the VWP to examine situations in which a speaker and addressee interact with a set of entities in a context in which only some of those entities are in common ground. Common ground is typically manipulated through either visual co-presence (whether an object can be seen by both partners), or linguistic co-presence (whether an object or piece of information has been mentioned in the context of both partners). Importantly, by creating situations in which some task-relevant objects are not in common ground—that is, they are in one partner's privileged ground—researchers can examine whether speakers and addressees take this perspective difference into account when producing or understanding the language of their partner (see Schober & Brennan, 2003). This ability to take a perspective difference into account is often described as 'perspective-taking' (cf. Baron-Cohen, Leslie, & Frith, 1985; Schober, 1993; Epley, Keysar, Van Boven, & Gilovich, 2004), however, it is important to note that in many cases, for

example the interpretation of an informational question, appropriate use of common ground information requires attention to the privileged ground. Thus, we define perspective-taking in language understanding not as the ability to adopt the perspective of one's conversational partner, but rather, the ability to appropriately attend to information that is either shared, or not shared with one's partner, depending on the context. In what follows, we review the key empirical findings regarding the role of common ground in incremental understanding, as well as the theories that have been developed to account for these findings.

3 Studies of On-line Perspective-taking

Empirical investigations of whether common ground guides on-line language processing have yielded results that appear to be contradictory. Some results show dramatic failures to use common ground when comprehending (Keysar, Lin, & Barr, 2003; Keysar, et al., 1998; Keysar, et al., 2000) or producing (Horton & Keysar, 1996) language, while others show that common ground is used only prior to receiving a critical linguistic stimulus, but not during its interpretation (Barr, 2008). These findings are contradicted by yet another group of results showing very early use of common ground (Hanna, et al., 2003; Heller, Grodner, & Tanenhaus, 2008; Brown-Schmidt, Gunlogson, & Tanenhaus, 2008; Brown-Schmidt, 2009a,b; Nadig & Sedivy, 2002; Metzing & Brennan, 2003). The emerging consensus is that common ground only sometimes influences language processing decisions, which leaves two key open questions: 1) when does the language processing system have access to information about perspective?; and 2) why do people appear to only sometimes show sensitivity to this information? With respect to the first question, the divisions among the primary candidate theories reflect earlier debates regarding the role of semantic and contextual information in syntactic parsing (e.g. Clifton & Ferreira, 1989; Tanenhaus, et al., 1995), with questions such as the timing of effects and resource-limitation issues taking center stage. With respect to the second question, these inconsistencies have been attributed to differences in experimental design characteristics, such as the type of ambiguity examined (Barr, 2008; Keysar, et al., 2003), or the way common ground was established (Hanna, et al., 2003), but other factors, such as culture (Wu & Keysar, 2007), inhibition control (Brown-Schmidt, 2009b; Nilsen & Graham, 2009), and mood (Converse, Lin, Keysar, & Epley, 2008; Van Berkum, 2009), are also thought to play a role.

There are three prominent candidate models of on-line perspective-taking, a Perspective-Adjustment model, an Anticipation-Integration model, and a Constraint-Based model. The critical difference between the competing theories lies in the proposed timing of perspective effects. Given recent improvements in methods for investigating the time-course of processing in contexts that afford perspective differences and the manipulation of perspective-taking, it would appear that an empirical resolution would be straightforward. In contrast, results from the most recent on-line research in this area are in conflict, and a comprehensive analysis of work in this area is lacking. As will become clear, these models are motivated by fundamentally distinct views of the nature of common ground, a by-product of which is critical differences in how common ground is experimentally manipulated. Consideration of such differences may allow a reconciliation of the contradictory findings.

In what follows, we outline the basic claims of the primary candidate theories, and present the evidence that is used to support them. Ultimately, we will argue that the bulk of this evidence is well accounted for within a Constraint-Based processing framework in which perspective is one of multiple, partial constraints on language processing. We will show that findings that apparently show failures or delays in the use of perspective actually highlight the way in which the language processing system is continuously sensitive to multiple, partial constraints, including perspective. In the process, we argue that careful consideration of how common ground is

established in experimental frameworks, and whether it is in conflict with other constraints, is critical to understanding how perspective information is used during language processing.

4 Perspective Adjustment

Perhaps the earliest proposal for how perspective information is incorporated into the incremental interpretation of utterances was the Perspective-Adjustment model (Keysar, et al., 1998; Keysar, et al., 2000; Keysar, et al., 2003; Keysar, 2007). This model was motivated by the assumptions that taking another person's perspective is resource intensive, and that most of the time perspective-taking is unnecessary because the perspectives of the speaker and addressee will overlap. The model proposed that the initial interpretation of a referring expression is unrestricted by common ground, only considering information available from the egocentric perspective of the addressee, and that a second monitoring process checks for violations of common ground, adjusting the initial interpretation as necessary (Keysar, et al., 1998).

Early support for this model came from studies showing that, when addressees interpreted their partners' referring expressions, private information (the 'privileged ground') competed for interpretation (Keysar, et al., 1998; Keysar, et al., 2000; Keysar, et al., 2003). The most prominent demonstration of what was termed an egocentric-first processing strategy comes from Keysar, et al. (2003). Participants in their first experiment followed instructions to manipulate objects, such as '*Pick up the tape*', in contexts that included a cassette tape in common ground and a roll of Scotch tape in privileged ground. During interpretation of *tape*, participants were five times more likely to fixate the privileged ground Scotch tape compared to a control condition where the privileged ground item was a battery. More strikingly, they reported that participants frequently reached for the privileged ground tape: 71% of participants attempted to reach for the privileged ground item on at least 1 out of 4 trials; 46% reached on at least 2 of 4 trials. This type of result is surprising in the context of the view that emphasizes common ground as a core component of meaning (e.g. Clark, 1992, 1996), and suggests that even in the ultimate interpretation of an expression, common ground does not always play a central role. What these results do not demonstrate, however, is what role common ground *does* have during interpretation.

Recent arguments in favor of the Perspective-Adjustment model have focused on the claim that taking another person's perspective is resource intensive. If so, this would suggest that it should be too difficult to routinely integrate perspective into initial language processing decisions. Evidence that perspective-taking is resource intensive comes primarily from measures and manipulations of working memory. For example, Lin, Keysar, and Epley (2010; also see Kronmüller & Barr, 2007, Horton & Keysar, 1996) used a design similar to Keysar, et al. (2003) and found that individuals with fewer working memory resources (due to lower working memory as measured in a span task, or due to an external load manipulation) looked at the competitor (e.g., *tape*) more than a control object (e.g., *battery*). Similarly, Epley, Morewedge, and Keysar (2004) found that children, who presumably have limited executive function, considered competitors more than adults. This evidence suggests that the resolution of competition in language processing is not always a resource-free process. What this evidence does not reveal, however, is whether perspective-taking is always a resource-intensive process, or whether perspective information is integrated into early on-line processing decisions, in spite of possible resource requirements. These issues are closely related to the question of what role, if any, common ground has in these experiments. We return to these issues in some detail in Section 7.

5 Anticipation Integration

More recently, Barr (2008) proposed a two-stage Anticipation-Integration account in which the language processing system has access to common ground in anticipation of (i.e., prior to) a referring expression, but when interpreting that expression, common ground does not play a role. According to this proposal, two distinct linguistic processes – anticipation and integration – show differential sensitivities to common ground. On this view, prior to hearing a referring expression (e.g., *the buckle*), an addressee might anticipate that her partner would refer to an object in common ground. Thus, unlike the Perspective-Adjustment model, common ground does routinely play a role in language processing, however this role is limited to the expectations that listeners form *prior* to the speaker's production of that expression. Much like the perspective-adjustment model, the anticipation-integration model proposes that *during* the processing of a referring expression, common ground is not integrated into the processing of that expression.

Support for this view comes from the results of three experiments in which addressees listened to instructions to manipulate various objects, some of which were in common ground with the speaker, and some of which were in the addressee's privileged ground. Critically, addressees tended to prefer to fixate common ground objects prior to hearing a referring expression, but during interpretation of that expression, looks to common and privileged objects increased at the same rate. For example, in Barr (2008) Experiment 1, participants followed pre-recorded instructions that they were led to believe were produced by a speaker in another room. Participants saw a screen with four pictures. A target picture (*buckle*) and two unrelated pictures were ostensibly shared with the speaker; a competitor picture was either shared or in the participant's privileged ground, and either had the same initial phonemes as the target (*bucket*), or did not (*ladder*). Prior to the onset of *buckle*, participants were significantly more likely to fixate the competitor when it was in common vs. privileged ground. This baseline difference between the two conditions was argued to be evidence for the use of common ground in the anticipation of an expression. Then, during interpretation of *buckle*, the increase in the number of fixations to the competitor bucket (i.e., the slope function relating fixation likelihood to time) was not affected by whether the competitor was in common or privileged ground. This result is consistent with the hypothesis that perspective does not guide interpretation of the noun. A second experiment which was explicitly non-interactive (subjects knew they were listening to recordings) showed weaker anticipatory effects. Again, during *buckle*, fixations to the competitor increased at a similar rate, regardless of ground. In a third experiment, participants were led to believe they were interacting with two other participants, a male and a female, from whom they heard pre-recorded instructions. Participants saw screens with three pictures, one of which was ostensibly in common ground with both speakers, one in common ground with the male speaker, and one in common ground with the female. Prior to the critical instruction, participants were more likely to look at a competitor if it was in common ground with the current speaker, however common ground played no role in the overall increase in fixations to the competitor during interpretation of the noun.

The results from these three experiments present a significant challenge to the view supported by Clark and his colleagues that common ground is a central component to language, defining the context within which comprehension takes place. Instead, they appear to suggest that language processing consists of at least two components, only one of which, anticipation, is sensitive to common ground. A key strength of the empirical findings is that the experiments do show effects of common ground during the anticipation period; this shows that participants were sensitive to the manipulation. However, a concern (that Barr raises) is whether these results are representative of processing in natural situations. After all, participants did not truly form common ground with other individuals. Instead, they were told what was shared with the two speakers, and then listened to pre-recorded instructions. A key question, then, is whether addressees would show

effects of common ground during incremental interpretation of language in more natural, interactive settings in which common ground is collaboratively established.

6 Constraint-Based Models

According to proponents of Constraint-Based models of perspective-taking in language processing (Nadig & Sedivy, 2002; Hanna, et al., 2003; Hanna & Tanenhaus, 2004; Heller, et al., 2008; Brown-Schmidt, et al., 2008; Brown-Schmidt, 2009b), perspective is only one of many partial constraints on language processing. Thus, on this view, evaluating whether an individual was sensitive to perspective when interpreting a referential ambiguity requires considering the other sources of information that may have influenced this process. How strongly a given constraint is predicted to bias interpretation in favor of one outcome over another depends critically on the relative strength of the evidence provided for the possible outcomes, and how strongly that constraint is weighted relative to other constraints. As an example, imagine a situation in which you and a friend go to the store, buy some Roquefort cheese, and jointly store it in the refrigerator. Unbeknownst to your friend, some fresh mozzarella is also in the fridge. If your friend were to later ask you *Can you grab some of that cheese for me?*, it would be clear that she meant the Roquefort, because only the Roquefort was established in common ground. In contrast, imagine your friend were simply reading from a recipe book, *The next ingredient is cheese*. In this case, which cheese you select from the fridge would be less influenced by the fact that the Roquefort is in common ground, than which type of cheese would go best with the dish. Thus, depending on characteristics of the discourse context and the linguistic expression, common ground may be more or less relevant to the understanding of language and associated behavioral goals.

Support for Constraint-Based accounts of perspective-taking comes from a growing number of findings that perspective information is used to reduce ambiguity. For example, participants in Hanna, et al. (2003)'s first experiment interpreted live instructions, produced by a confederate speaker, such as *Put the blue triangle on the red one*, in contexts that contained one blue triangle and two red triangles. The critical manipulation was whether both red triangles were in common ground, or if one was in common ground and the other was in the addressee's privileged ground. In this situation, a variety of constraints are relevant to interpretation of the critical noun phrase, *the red one*. The most relevant constraints are common ground and lexical information: The use of the definite determiner *the* indicates that the referent should be in common ground (or uniquely identifiable in the discourse context, see Roberts, 1993), and the lexical information in *red* indicates that the referent should be red. The Constraint-Based view predicts that these constraints should combine to influence identification of the referent, thus there should be competition between the red triangles that is diminished somewhat when only one is in common ground.

The results were consistent with these predictions: When both red triangles were in common ground, during the interpretation of *red*, participants showed a significant competition effect, with equivalent fixations to the two red triangles. In contrast, when one of the red triangles was in privileged ground, participants primarily looked at the common ground triangle within the first few hundred milliseconds of processing. Critically, however, as predicted by Constraint-Based processing, there was still a lexical competition effect; participants were significantly more likely to fixate the privileged ground competitor triangle when it was red compared to a condition in which it was yellow. Hanna et al. concluded that common ground acts immediately, but probabilistically, to restrict the domain of interpretation; this means that objects that are visually available and salient, and that match the referring expression, can interfere with reference resolution even if they are in privileged ground, but that common ground acts immediately to probabilistically restrict the addressee's domain of interpretation to those objects that can be referred to by the speaker. This observation of the simultaneous effects of many constraints is consistent with a large body of work showing that multiple, sometimes competing sources of

information combine to influence on-line processing (Novick, et al., 2008; Britt, 1994; Snedeker & Yuan, 2008; Hanna & Tanenhaus, 2004; Garnsey, Pearlmuter, Myers, & Lotocky, 1997; Wilson & Garnsey, 2009).

One objection to these results was that participants might have been cued to use perspective information due to the fact that the instructions would have been globally ambiguous without taking perspective into account (Barr, 2008; also see Keysar, et al., 2003). The concern, then, was that the results do not reflect typical processing, based on an assumption that global ambiguities are infrequent in natural language. However, the results have since been replicated with adults in constructions with temporary ambiguities, i.e. ambiguities that are resolved linguistically before the end of the phrase (Heller, et al., 2008; also see Brown-Schmidt, et al., 2008; Brown-Schmidt, 2009b), thus the conclusion that participants used perspective information to constrain interpretation of these expressions seem warranted. Further, results similar to Hanna, et al.'s have also been observed in young children (Nadig & Sedivy, 2002; Nilsen & Graham, 2009; Scott, Brown-Schmidt, Fisher, & Baillargeon, 2009), and in individuals with amnesia (Rubin, Brown-Schmidt, Duff, Tranel, & Cohen et al., 2009), suggesting that use of common ground may not be exceptionally resource-intensive, contrary to some claims (Converse, et al., 2008; Keysar, 2007; Lin, et al., 2010).

Sometimes perspective-taking actually requires attending to privileged information rather than ignoring it. Consider the case of an informational question. A felicitous informational question typically asks about information in the addressee's privileged ground. That is, speakers tend to ask questions when they don't know the answer but they believe that the addressee might. Thus from the addressee's perspective, a perspective-appropriate interpretation of what an informational question is asking about would be information in the addressee's privileged ground. In previous work, perspective-taking has typically been defined as interpreting imperatives or indirect requests as referring to common ground referents. In contrast, Brown-Schmidt and colleagues (Brown-Schmidt, et al., 2008; Brown-Schmidt 2009b) have demonstrated use of perspective information during the on-line interpretation of informational wh-questions. For example, Brown-Schmidt (2009b) asked participants *What's above the cow that's wearing shoes?*, in contexts that contained two cows (one with shoes, one with glasses). The critical manipulation was whether the animal above the competitor cow (with glasses) was in common ground. Following *cow*, participants' preference to fixate the privileged ground target was significantly higher when the animal above the competitor was in common ground. Thus, addressees showed sensitivity to perspective by directing attention towards *privileged* ground entities when interpreting temporarily ambiguous questions. This finding places doubt on claims that perspective-taking involves serial adjustment away from the egocentric perspective (i.e., Epley, Keysar, et al., 2004), as in some cases, perspective-taking actually requires attending to privileged information.

7 Reconciling Previous Findings

In this section, we will show that each of the major findings on the time-course of perspective-taking during language processing can be accounted for under the Constraint-Based processing framework. In contrast, the major alternative theories, Perspective-Adjustment and Anticipation-Integration, fail to account for key findings that support the Constraint-Based view.

Let us first consider findings by Keysar, et al.'s (2003) experiment, that participants were more likely to fixate a privileged-ground competitor when it matched the referential description *tape* than when it did not (i.e., Scotch tape vs. battery). While this result was interpreted as evidence against the use of perspective information, according to the Constraint-Based view, this result shows nothing more than evidence of lexical competition effects in interpretation. Participants should always fixate a referent that matches the critical referring expression more

than one that does not, since lexical information is a very good cue to speaker meaning. The extent of interference from privileged ground competitors in this experiment was further amplified by a meaning dominance effect: Each of the competitor objects (e.g., Scotch tape) were explicitly designed to be a better match to the critical referring expression, *tape*, than the common ground target (e.g., cassette tape). Similar problems are present in several other studies that purport to show evidence for difficulties in perspective-taking (e.g., Converse, et al., 2008; Epley, et al., 2004; Keysar, et al., 1998; Keysar, et al., 2000; Lin, et al., 2010; Wu & Keysar, 2007). For example, in Keysar, et al. (2000), the director referred to *the bottom block* in contexts with three blocks, two in common ground and one in privileged ground, each on a different row of the display. However, in the critical conditions, the privileged object was always the best perceptual match for the referring expression. In this example, the block in privileged ground was always the bottom-most block, and therefore matched the referring expression, *the bottom block*, the best. Similar to the Keysar, et al. (2003) results, participants were not only just as likely to look at the blocks in privileged and common ground, but they actually initially preferred to look at the block in privileged ground, a result that Keysar, et al. (2003) used to support their view that initial interpretation is egocentric. Under these circumstances, it is difficult to know whether common ground also played a role, since this result only shows that common ground did not completely rule out consideration of the privileged ground object when there was lexical competition between two interpretations, and a lexical bias in favor of a privileged competitor.

What role might perspective information have in these situations? According to Constraint-Based models, perspective information would have offered some support to the common-ground interpretation of the critical expression. Critically, however, the experiments by Keysar, et al. (2003) and Keysar, et al. (2000) lacked the crucial comparison condition that would have revealed the common ground effect—one in which the critical competitor is in *common* ground (e.g., the Scotch tape or the bottom-most block). This key condition would likely have shown that participants are less likely to fixate a competitor when it is privileged versus common (i.e., the common ground effect observed in Hanna, et al., 2003). Thus, experimental designs such as the one used by Keysar, et al. (2003; also Converse, et al., 2008; Epley, Morewedge, et al., 2004; Keysar, et al., 1998; Keysar, et al., 2000; Lin, et al., 2010; Wu & Keysar, 2007) are not designed in such a way that perspective effects could be revealed, even if they were to occur.

Similar problems confound the interpretation of findings that individuals with fewer resources—individuals with low, or taxed working memory (Lin, et al., 2010), or children (Epley, Morewedge, et al., 2004) were more likely to gaze at a privileged ground object when it was a strong lexical competitor, than when it was not. While these results were taken as evidence that *perspective-taking* requires mental effort, in our view, what they actually show is that resolving *lexical competition* requires resources. The crucial comparison that would test the use of perspective in the face of resource limitations—a condition in which both the target and lexical competitor were common ground—was missing. Several other concerns limit the impact of these findings. One is the assumption that if perspective-taking is resource-intensive, then it cannot guide on-line processing. This has not been demonstrated experimentally, and the fact that many studies have now shown successful use of common ground on-line with adults (Hanna, et al., 2003; Heller, et al., 2008; Brown-Schmidt, et al., 2008; Brown-Schmidt, 2009b), and even children (Nadig & Sedivy, 2002; Nilsen & Graham, 2009; Matthews, Lieven, & Tomasello, 2010), suggests that even if the premise is true, the conclusion is false. A further consideration is that measures and manipulations of working memory (also Kronmüller & Barr, 2007; Horton & Keysar, 1996) may be more informative about language experience than capacity (see MacDonald & Christiansen, 2002; Wells, et al., 2009). Finally, in some cases computing and accessing representations of common ground may be highly automatized, due to associations in memory between partners and shared information (Horton & Gerrig, 2005a,b; Horton, 2007), or low-level cues to joint knowledge, such as shared gaze (Hanna & Brennan, 2007).

Let us now consider the studies reported in Barr (2008) which showed evidence of perspective-taking in anticipation of hearing a critical expression, but not during its interpretation. Unlike the studies by Keysar and colleagues, these experiments did include the critical comparison condition in which the competitor was in common ground. The conclusion drawn from the results of these experiments, that perspective does not affect on-line interpretation of the critical expressions, is incompatible with the Constraint-Based prediction. Further, the fact that there was a significant perspective effect prior to the critical expressions suggests that this was not simply a null effect due to a weak manipulation of perspective. However, there are several reasons to treat these results with caution.

Our first objection concerns the claim that there are two separate processing stages—anticipation and integration—and the finding that only the former is sensitive to perspective. If the baseline preference to fixate common ground entities were truly an anticipatory effect, it would suggest that the processing system always anticipates common ground referents prior to a critical expression. However, this can't possibly hold for normal conversation, because, as Brown-Schmidt and colleagues (2008) pointed out, whether shared or private information is relevant depends on utterance form (e.g. whether it is a statement or a question). A processing mechanism that focused attention on common ground information prior to each linguistic act would frequently make the wrong prediction. Indeed, in the only study to examine perspective-taking in completely unscripted conversation between naïve partners (Brown-Schmidt, et al., 2008, Experiment 2), addressees were equally likely to gaze at common and privileged entities prior to the onset of critical expressions. Given this, it would seem that the ‘anticipatory’ effects may simply reflect processing of the initial words in the speaker’s request; in Barr’s first experiment, they were *Click on the*. As participants heard these words, they were interpreted as a request to perform an action on some object, with the constraint that it should be a shared object.

The second objection has to do with the finding that perspective played no role in interpretation of the critical word *bucket*, which was temporarily ambiguous between the shared target and the privileged competitor, ‘buckle’. According to Constraint-Based theories, at least two constraints should play a role in the interpretation: perspective, which should have supported the ‘bucket’ interpretation; and lexical information, which should have temporarily supported both ‘buckle’ and ‘bucket’ interpretations. Why, then, did fixations to buckle and bucket increase at the same rate? We suspect the answer may lie in the fact that participants did not actually form common ground in this experiment. Recall that according to classic accounts, common ground forms as individuals *collaboratively* establish what information is jointly known through an interactive grounding process (Brennan & Clark, 1996). In each of the studies that have shown significant effects of common ground in on-line interpretation, participants interacted with live partners with whom they were able to collaboratively form common ground (e.g., Hanna, et al., 2003; Nadig & Sedivy, 2002; Heller, et al., 2008; Brown-Schmidt, et al., 2008; Brown-Schmidt, 2009a,b; Metzing & Brennan, 2003). In contrast, in Barr’s (2008) experiments, participants never interacted with live partners, and never engaged in grounding procedures. Instead, in the first experiment, a marking on a card indicated which object was shared with the speaker. In the second experiment, participants knew they were listening to pre-recorded instructions and again saw markings on cards to cue which object was shared. Finally, in the third experiment, participants listened to recordings from what they were led to believe were two different speakers. This experiment contained a pseudo-grounding procedure in which the participant overheard the experimenter and one of the pre-recorded voices ‘discuss’ what was in common ground. This design would be expected to produce weak common ground effects as it is well known that overhears do not form common ground in the same way as full participants (Schober & Clark, 1989; also see Wilkes-Gibbs & Clark, 1992). Consistent with this assertion is recent evidence from Brown-Schmidt (2009a) that addressees show immediate sensitivity to joint knowledge in interactive settings only.

Re-framing these findings in such a way, it becomes clear that the Constraint-Based framework would predict the observed findings given the weak manipulation of common ground in conjunction with a powerful lexical competition effect: Prior to the critical noun, participants incrementally interpreted the utterance *Click on the...* as referring to shared objects, due to the absence of competing constraints and a weak perspective effect. Then, during interpretation of the critical noun *buckle*, lexical constraints were in conflict with a relatively weak perspective effect. These lexical constraints overwhelmed perspective information, resulting in an equivalent increase in fixations to the target and competitor¹. Because of the weak perspective manipulation, then, the experiment was not designed in such a way to distinguish between Anticipation-Integration and Constraint-Based accounts.

A final consideration is that individuals may vary in their sensitivity to perspective information, or the degree to which they are able to successfully combine competing sources of information to arrive at an appropriate interpretation of the speaker's meaning. Taking into account such individual differences may provide a partial explanation for the large amount of variability in perspective-taking findings, and for why some participants fail to use perspective information in the face of competing constraints. For example, inhibition control, which is known to predict children's performance in theory-of-mind tasks (Carlson & Moses, 2001; Chasiotis, Kiessling, Hofer, & Campos, et al., 2006; Hughes & Ensor, 2005) also predicts the use of perspective information during on-line language processing in both children (Nilsen & Graham, 2009), and adults (Brown-Schmidt, 2009b), with individuals who score higher on conflict inhibition tasks showing better ability to rule out perspective-inappropriate interpretations of their partner's speech. The locus of the effect is unknown, but may be related to improved inhibition of the perspective-inappropriate interpretation. Alternatively, individuals with higher inhibition control may simply be more skilled at resolving conflicts between multiple constraints (also see Novick, et al., 2008). Mood and culture may also play a role. For example, Converse, et al., (2008) found that participants in a happy mood were more likely to adopt perspective-inappropriate interpretations, and Wu & Keysar (2007) found that Chinese participants were less likely to adopt perspective-inappropriate interpretations compared to their American counterparts. On the Constraint-Based view, the fact that factors such as these modulate whether an addressee will entertain a perspective-inappropriate response suggests that it is unlikely that there exists a perspective-free processing stage, but instead that the likelihood that perspective will guide interpretation is dependent on a variety of factors.

8 Evaluating Constraint-Based Models

In this section, we will consider what sorts of experimental manipulations would provide critical tests of the Constraint-Based model. The Constraint-Based approach proposes that perspective

¹ It is worth noting that there are additional concerns regarding the statistical support for these anticipation and integration effects. Barr (2008) used a regression approach, and equated anticipatory effects with condition differences at the intercept, and integration effects with slope differences, that is, differences between the conditions in the change of the likelihood of a target fixation over time. This statistical approach of modeling intercept effects separately from slope effects has been questioned (Tanenhaus, et al., 2008), as slopes and intercepts are likely to be non-independent (even in log-odds space). The independence assumption requires the population of trials on which there was not a baseline target fixation to be representative and comparable across the conditions being compared, as the rise in slope following the critical noun is largely driven by shifts in fixations from non-targets to the target. However, if only a subset of the trials (or participants) are likely to be affected by the perspective manipulation (e.g., due to individual differences in mood [Converse, et al., 2008], inhibition control [Brown-Schmidt, 2009b], etc.), those datapoints would be systematically excluded from the slope measure, as they would have already shown sensitivity to the manipulation at baseline. An alternative is to eliminate baseline (intercept) effects by analyzing the sub-set of trials without a baseline target fixation, however this approach can result in significant data loss (e.g., Brown-Schmidt, 2009b). Alternative designs, such as those in which attention is drawn to an unrelated object at baseline (see Hanna, et al., 2003) may be required to resolve these issues.

information, along with multiple other partial constraints, combine to influence interpretation processes. A growing number of results are consistent with the Constraint-Based view, including findings that addressees use information about the perspective of their partner to guide on-line understanding (Nadig & Sedivy, 2002; Hanna, et al., 2003; Hanna & Tanenhaus, 2004; Heller, et al., 2008; Brown-Schmidt, et al., 2008; Brown-Schmidt, 2009b; or character in a narrative, Ferguson, Scheepers, & Sanford, 2010), as well as findings that addressees are sensitive to partner-specific referring history when interpreting referring expressions (Brown-Schmidt, 2009a, Metzing & Brennan, 2003; Brennan & Hanna, 2009; Matthews, et al., 2010). In the previous section, we have shown that the Constraint-Based model of perspective-taking can account for the wide range of findings that were once thought to be inconsistent with the Constraint-Based approach. As we demonstrated, considerations of how perspective was established, and what other constraints were relevant, resolve such inconsistencies.

In the past, implemented constraint based models have been used to generate testable predictions about the strength of verb bias, thematic fit, and discourse context effects, and the subsequent predictions for interpretation of attachment ambiguities (Spivey & Tanenhaus, 1998; McRae, Spivey-Knowlton, & Tanenhaus, 1998; Tanenhaus, Spivey-Knowlton, & Hanna, 2000). A similar extension to common ground, however, proves more difficult, because quantifying the degree to which common ground picks out a particular referent is not straightforward. A more tangible goal may be to manipulate the relative strength of common ground, an approach that was successfully applied by Brown-Schmidt (2009a) to the question of whether collaboratively-defined expressions are part of partner-specific common ground. A similar approach could be taken to examine the role of perspective information in on-line processing. Generally speaking, the Constraint-Based view predicts that the strength of perspective representations should directly modulate their effectiveness, with the strongest effects of common ground predicted when it is established interactively. If a weak common ground manipulation were pitted against a strong competing constraint (e.g., see Barr, 2008), the perspective effect is expected to be small or possibly eliminated. For example, if participants were instructed to *Put the star below the bucket* in a context that included a common-ground ‘bucket’, and a ‘buckle’ that was either common or privileged, the Constraint-Based account would predict that during interpretation of *bucket*, participants should experience less competition from the buckle when it was privileged, compared to common. However, the magnitude of this competition reduction should be directly modulated by the strength of the common ground manipulation, with larger reductions in interactive conversation compared to non-interactive settings.

A perhaps unsatisfying implication of such predictions is that in cases where the Constraint-Based view predicts significant effects of ground and interactivity, the competing models predict null effects during on-line interpretation. Another way to tease apart these models would be to examine situations in which the theories make opposite predictions. A candidate test could involve examination of spatial perspective-taking: Consider the fact that in a face-to-face conversation, one’s egocentric ‘left’ is one’s partner’s ‘right’. In face-to-face task-based conversation, dialog partners often negotiate such terms by agreeing to re-define these terms from a particular perspective (although sometimes neutral perspectives are adopted, or a more skilled partner accommodates to a less skilled partner, c.f., Schober, 1993; 1995). For example, consider the following exchange, from naïve participants on either side of a 3-d display (Brown-Schmidt, et al., 2008, Experiment 2):

1. *yeah what's to the left of that?*
2. *ahhh*
1. *or to the right?*
2. *pig with sneakers... it's always gonna be from your perspective pig with sneakers*

Now consider a similar situation in which partners play a game in which they sit face-to-face on either side of a display with animals, and direct each other to re-arrange shapes in the display (see Figure 1). Imagine that the partners jointly define “right” from Partner 1’s perspective, thus making this new definition common ground. If Partner 1 were to subsequently say *Take the star and put it below the dog on the right*, in a context containing two dogs, one on the left side of the display, and one on the right, it would be possible to examine Partner 2’s incremental interpretation of the expression. According to Constraint-Based models, at *right*, Partner 2 should immediately access the collaboratively-defined meaning of *right* as defined from Partner 1’s perspective, and interpret the expression as referring to the dog wearing the purse. Lexical competition from the previous word *dog*, and the overall frequency in the language of using *right* to refer to the egocentric right would be competing constraints, partially supporting interpretation of the expression as the dog wearing the flower. According to the Constraint-Based view, these competing constraints would manifest as increased fixations to the competitor dog (wearing the flower) compared to the cows in the scene.

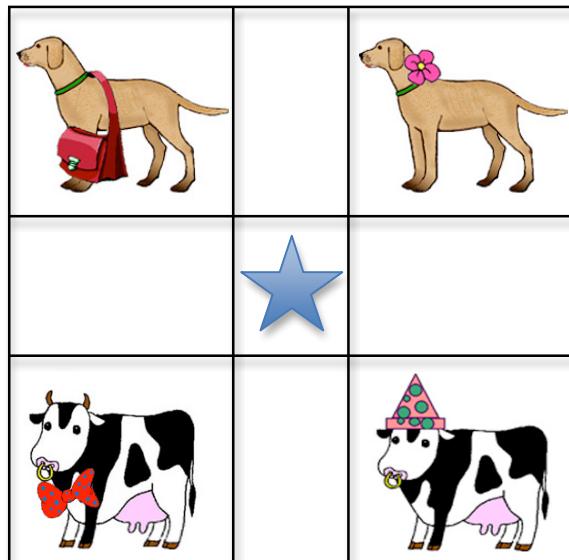


Figure 1. Example display from partner 2’s perspective. Partner 1 would see the mirror-image reverse.

In contrast, on the Perspective-Adjustment model, when partners have opposite perspectives, the initial interpretation of *right* should be egocentric, as the dog wearing the flower. Similarly, on the Anticipation-Integration account, during the on-line interpretation (integration) of *right*, perspective should not be relevant, thus the expression should be interpreted as referring to the dog wearing the flower. A crucial test of the Constraint-Based account would come from a case in which partner 1 violated common ground, and used the spatial term from partner 2’s perspective: *Take the star and put it below the dog on the right wearing the flower*. On Perspective-Adjustment and Anticipation-Integration accounts, this utterance is consistent with the egocentric perspective, and should be easy to interpret. In contrast, on the Constraint-Based account, this sentence should create a garden-path effect, and should be confusing. A comparison condition would come from a case in which the partners jointly agreed to define *right* as from partner 2’s perspective. According to the Constraint-Based account, the critical sentence should be significantly easier to understand when *right* was defined from 2’s perspective. There should be no difference according to the competing theories, since both would predict that interpretation of *right* from partner 2’s perspective should be relatively easy.

Finally, while empirical investigations such as the one we suggested will be important as we evaluate and improve models of on-line processing, it is important to point out that much of the research and theorizing about the on-line use of perspective has focused on issues of timing. One consequence of this near-singular focus on the timing of perspective effects is that little is known about the nature of perspective representations themselves. Attention to such matters, however, is necessary in order to build explicit models of perspective-taking, and may be important for fleshing out predictions regarding the difficulty of rapidly engaging perspective representations during on-line interpretation. There are at least two views of how perspective representations might be encoded. Early psychological views of common ground postulated that interlocutors form rich representations of each other's knowledge that go well beyond whether a given object is visually co-present or not. For example, Clark and Marshall (1978, 1981) proposed that interlocutors store rich, diary-like information about events and the people they experienced those events with. They argued that information enters the common ground through a variety of different routes including not only what is jointly visible, but also what has been said, what can be jointly inferred, and what is general knowledge in the community. Speakers and addressees alike access these representations of jointly experienced events, dialogs, and community knowledge and use this information to support the use of language in conversational settings. Whether representations as rich as the ones that Clark and Marshall postulated are actually used on-line in conversation is a topic of current debate. While there has been some study of the use of these different types of common ground in language production and in the ultimate understanding of an utterance (Fussell & Krauss, 1991, 1992; Issacs & Clark, 1987; also see Schober & Brennan, 2003), most if not all of the existing research on the on-line use of common ground has focused on simple visual or linguistic co-presence (e.g. Keysar, et al., 1998; Keysar, et al., 2003; Hanna, et al., 2003; Brown-Schmidt, et al., 2008; Nadir & Sedivy, 2002; Brown-Schmidt, 2009b). Another possibility, suggested by Hanna and colleagues (Hanna, et al., 2003; also see Horton, 2007), is that sensitivity to common ground may be largely supported by low-level information sources such as eye-gaze (see Hanna & Brennan, 2007; Richardson, Dale, & Kirkham, 2007), gesture and body position, and possibly even very basic co-occurrence information (Horton & Gerri, 2005a, b; Horton, 2007). Understanding whether the use of common ground requires accessing rich, event-based representations (Clark & Marshall, 1978, 1981), or sensitivity to low-level cues (Hanna & Brennan, 2007; Horton & Gerri, 2005a, b), or both, will likely provide crucial insights into the mechanisms by which this information is integrated into the on-line processing of utterances.

9 Conclusions

In this article, we focused on one type of contextual information, specifically the common-ground status of the entities in the discourse context. Like earlier debates about the role of context in syntactic processing decisions, the crucial theoretical question was not whether common ground played a role in processing, but when, during the time-course of understanding an utterance, common ground was relevant. A second important question that has emerged is why there is so much variability between individuals and across contexts in the use of common ground. As we have discussed, the candidate theories—the Perspective-Adjustment, Anticipation-Integration, and Constraint-Based models—primarily differ in terms of when, in time, common ground is thought to guide interpretation of referring expressions in dialog. Thus, like earlier debates about the role of context in syntactic ambiguity resolution, the precise time-scale of eye-tracking data in combination with the ability to monitor eye movements in rich discourse contexts has made the VWP the dominant methodology for examining the role of common ground in on-line language understanding. The now large number of studies using this paradigm have generated many insights, and puzzles, about how common ground information guides on-line language understanding.

Here we have argued for a Constraint-Based model of the role of common ground in on-line language understanding. We proposed that common ground, or more generally, perspective information, is always potentially available to language processing decisions, but that the likelihood an addressee will adopt a perspective-appropriate interpretation of her partner's utterance will depend on the strength and relevance of the perspective representation, and whether perspective is in conflict with other constraints. This view can account for a large number of results in the literature, including findings that as an addressee follows instructions from a dialog partner, she is likely to consider all entities that match her partner's referring expression (Keysar, et al., 1998; Keysar, et al., 2000; Keysar, et al., 2003), with a preference for those in common ground (Hanna, et al., 2003, Nadir & Sedivy, 2002). This view is also consistent with findings that as an addressee interprets an informational wh-question, she is likely to interpret that question as asking about information only the addressee knows about (Brown-Schmidt, et al., 2008; Brown-Schmidt, 2009b). While perspective is always available, in some circumstances it may be less relevant to language processing decisions. For example, if the addressee is listening to pre-recorded speech (e.g. Barr, 2008), as in the pre-recorded messages played in airports, or if the speaker is not herself the author of the ideas, as when reading a recipe aloud or giving a survey (see Schober & Conrad, 2008), the addressee may rightly infer that the speaker's knowledge state is of little relevance to understanding what is being said. In these circumstances, according to the Constraint-Based model, the addressee would be expected to weight other sources of information, such as lexical meaning or thematic fit, more heavily (see Brown-Schmidt, 2009a).

In conclusion, using language in conversation requires speakers and listeners to quickly incorporate many different sources of information into language production and comprehension processes. Here, we focused primarily on the task of the addressee. As she listens to her partner's speech, she must rapidly and incrementally interpret each word and utterance, making many provisional commitments about the intended meaning of each word or structure. As has been clearly articulated before, we arrive at meaning in language not through the linguistic stimulus alone, but by combining it with information gleaned from our background knowledge and a rich discourse context (Clark, 1992, 1996). For scholars of language processing, the domain of inquiry, then, is to examine what sources of information are relevant, and the mechanisms by which they are integrated into our understanding of language.

Acknowledgements

This material is based partially upon work supported by the National Science Foundation under Grant No. BCS-10-19161 to the first author, and Grant No. IIS-07-13287 to the second author. The authors would like to thank three anonymous reviewers for their insightful comments and suggestions. J. E. Hanna would like to thank Timothy Dustin, Sarah Mundell, Barbara Percival, and Anne Wildman for their contributions. Some of the material in this paper was presented by J.E. Hanna at the Workshop on Incrementality in Verbal Interaction (June, 2009) at ZiF (Centre for Interdisciplinary Research), University of Bielefeld, Germany.

References

- Gerry T. M. Altmann and Yuki Kamide (1999). Incremental interpretation at verbs: restricting the domain of subsequent reference. *Cognition*, 73:247-264.
- Gerry T. M. Altmann and Yuki Kamide (2007). The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *Journal of Memory and Language*, 57:502-518.

- Gerry T. M. Altmann and Yuki Kamide (2009). Discourse-mediation of the mapping between language and the visual world: Eye movements and mental representation. *Cognition*, 111:55-71.
- Paul D. Allopenna, James S. Magnuson, and Michael K. Tanenhaus (1998). Tracking the time course of spoken word recognition: evidence for continuous mapping models. *Journal of Memory and Language*, 38:419-439.
- Jennifer E. Arnold and Zenzi M. Griffin (2007). The effect of additional characters on choice of referring expression: Everyone counts. *Journal of Memory and Language*, 56:521-536.
- Jennifer E. Arnold, Carla L. Hudson Kam, and Michael K. Tanenhaus (2007). If you say *Thee uh* you are describing something hard: The on-line attribution of disfluency during reference comprehension. *Journal of Experimental Psychology, Learning, Memory and Cognition*, 33:914-930.
- Simon Baron-Cohen, Alan M. Leslie and Uta Frith (1985). Does the autistic child have a “theory of mind”? *Cognition*, 21:37-46.
- Dale J. Barr (2008). Pragmatic expectations at linguistic evidence: Listeners anticipate but do not integrate common ground. *Cognition*, 109:18-40.
- Robbert-Jan Beun and Anita H. M. Cremers (1998). Object reference in a shared domain of conversation. *Pragmatics & Cognition*, 6:121-151.
- Susan E. Brennan and Herbert H. Clark (1996). Conceptual Pacts and Lexical Choice in Conversation. *Journal of Experimental Psychology: Learning, Memory & Cognition*, 22:482-493.
- M. Anne Britt (1994). The interaction of referential ambiguity and argument structure in the parsing of prepositional phrases. *Journal of Memory and Language*, 33:251-283.
- Sarah Brown-Schmidt (2009a). Partner-specific interpretation of maintained referential precedents during interactive dialog. *Journal of Memory and Language*, 61:171-190.
- Sarah Brown-Schmidt (2009b). The role of executive function in perspective-taking during on-line language comprehension. *Psychonomic Bulletin and Review*, 16:893-900.
- Sarah Brown-Schmidt and Michael K. Tanenhaus (2008). Real-time investigation of referential domains in unscripted conversation: A targeted language game approach. *Cognitive Science*, 32:643-684.
- Sarah Brown-Schmidt, Christine Gunlogson, and Michael K. Tanenhaus (2008). Addressees distinguish shared from private information when interpreting questions during interactive conversation. *Cognition*, 107:1122-1134.
- Sarah Brown-Schmidt, Ellen Campana, and Michael K. Tanenhaus (2005). Real-time reference resolution by naïve participants during a task-based unscripted conversation. *Approaches to studying world-situated language use: Bridging the language as product and language as action traditions*, Eds. John C. Trueswell and Michael K. Tanenhaus, pages 153-171. MIT Press, Cambridge, Massachusetts.
- Stephanie M. Carlson and Louis J. Moses (2001). Individual differences in inhibitory control and children’s theory of mind. *Child Development*, 72:1032–1053.
- Athanasiос Chasiotis, Florian Kiessling, Jan Hofer, and Domingo Campos (2006). Theory of mind and inhibitory control in three cultures: Conflict inhibition predicts false belief understanding in Germany, Costa Rica and Cameroon. *International Journal of Behavioral Development*, 30:249-260.
- Craig G. Chambers, Michael K. Tanenhaus, Kathleen M. Eberhard, Hana Filip, and Greg N. Carlson (2002). Circumscribing Referential Domains during Real-Time Sentence Comprehension. *Journal of Memory and Language*, 47:30-49.
- Craig G. Chambers, Michael K. Tanenhaus, and James S. Magnuson (2004). Actions and Affordances in Syntactic Ambiguity Resolution. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 30:687-696.

- Herbert H. Clark (1992). *Arenas of Language Use*. University of Chicago Press, Chicago, Illinois.
- Herbert H. Clark (1996). *Using Language*. Cambridge University Press, Cambridge, United Kingdom.
- Herbert H. Clark and Catherine Marshall (1978). Reference diaries. *Theoretical issues in natural language processing* (Vol. 2), ed. David L. Waltz, pages 57-63. Association for Computing Machinery, New York, New York.
- Herbert H. Clark and Catherine Marshall (1981). Definite reference and mutual knowledge. *Elements of Discourse Understanding*, eds. Aravind K. Joshi, Bonnie L. Webber, Ivan A. Sag, pages 10-63. Cambridge University Press, Cambridge, United Kingdom.
- Herbert H. Clark and Deanna Wilkes-Gibbs (1986). Referring as a collaborative process. *Cognition*, 22:1-39.
- Charles Clifton, Jr. and Fernanda Ferreira (1989). Ambiguity in Context. *Language and Cognitive Processes*, 4:77-103.
- Benjamin A. Converse, Shuhong Lin, Boaz Keysar, and Nicholas Epley (2008). In the mood to get over yourself: Mood affects theory-of-mind use. *Emotion*, 8(5):725-730.
- Roger M. Cooper (1974). The control of eye fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology*, 6:84-107.
- Delphine Dahan, James S. Magnuson, Michael K. Tanenhaus, and Ellen M. Hogan (2001). Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition. *Language and Cognitive Processes*, 16(5/6):507-534.
- Delphine Dahan and Michael K. Tanenhaus (2004). Continuous Mapping from Sound to Meaning in Spoken-Language Comprehension: Immediate Effects of Verb-Based Thematic Constraints. *Journal of Experimental Psychology, Learning, Memory and Cognition*, 30:498-513.
- Delphine Dahan and Michael K. Tanenhaus (2005). Looking at the rope when looking for the snake: Conceptually mediated eye movements during spoken-word recognition. *Psychonomic Bulletin & Review*, 12:453-459.
- Dieuwke De Goede, Petra Van Alphen, Emma Mulder, Yvonne Blokland, José Kerstholt, and Jos J. A. Van Berkum (2009). The effect of mood on anticipation in language comprehension: An ERP study. Poster presented at the 16th Annual meeting of the Cognitive Neuroscience Society (CNS-2009), San Francisco.
- Kathleen M. Eberhard, Michael J. Spivey-Knowlton, Julie C. Sedivy, and Michael K. Tanenhaus (1995). Eye-movements as a window into spoken language comprehension in natural contexts. *Journal of Psycholinguistic Research*, 24:409-436.
- Nicholas Epley, Boaz Keysar, Leaf Van Boven, and Thomas Gilovich (2004). Perspective taking as egocentric anchoring and adjustment. *Journal of Personality and Social Psychology*, 87(3):327-339.
- Nicholas Epley, Carey K. Morewedge, and Boaz Keysar (2004). Perspective taking in children and adults: Equivalent egocentrism but differential correction. *Journal of Experimental Social Psychology*, 40:760-768.
- Heather J. Ferguson, Christoph Scheepers, and Anthony J. Sanford (2010). Expectations in counterfactual and theory of mind reasoning. *Language and Cognitive Processes*, 25:297-346.
- Fernanda Ferreira and Charles C. Clifton, Jr. (1986). The independence of syntactic processing. *Journal of Memory and Language*, 25:348-368.
- Susan R. Fussell and Robert M. Krauss (1991). Accuracy and bias in estimates of others' knowledge. *European Journal of Social Psychology*, 21:445-454.
- Susan R. Fussell and Robert M. Krauss (1992). Coordination of Knowledge in Communication: Effects of Speakers' Assumptions About What Others Know. *Journal of Personality and*

- Social Psychology*, 62:378-391.
- Susan M. Garnsey, Neal J. Pearlmuter, Elizabeth Meyers, and Melanie A. Lotocky (1997). The contributions of verb bias and plausibility to the comprehension of temporarily ambiguous sentences. *Journal of Memory and Language*, 37:58-93.
- Zenzi M. Griffin and Kathryn Bock (2000). What the eyes say about speaking. *Psychological Science*, 11(4):274-279.
- Herbert P. Grice (1975). Logic and Conversation. *Syntax and semantics 3: Speech Arts*, eds. Peter Cole and Jerry L. Morgan, pages 41-58. Academic Press, New York, New York.
- Daniel Grodner and Julie Sedivy (in press). The effects of speaker-specific information on pragmatic inferences. *The Processing and Acquisition of Reference*, eds. N. Pearlmuter & E. Gibson. MIT Press: Cambridge, Massachusetts.
- Peter E. Hallett (1986). Eye movements. *Handbook of perception and human performance (Vol. 1)*, eds Kenneth R. Boff, Lloyd Kaufman, and James P. Thomas, pages 10.1-10.112. Wiley, New York, New York.
- Joy E. Hanna and Susan E. Brennan (2007). Speakers' eye gaze disambiguates referring expressions early during face-to-face conversation. *Journal of Memory and Language*, 57:596-615.
- Joy E. Hanna, Michael K. Tanenhaus, and John C. Trueswell (2003). The effects of common ground and perspective on domains of referential interpretation. *Journal of Memory and Language*, 49:43-61.
- Joy E. Hanna and Michael K. Tanenhaus (2004). Pragmatic effects on reference resolution in a collaborative task: evidence from eye movements. *Cognitive Science*, 28:105-115.
- Daphna Heller, Daniel Grodner, and Michael K. Tanenhaus (2008). The role of perspective in identifying domains of reference. *Cognition*, 108:831-836.
- William S. Horton (2007). The influence of partner-specific memory associations on language production: Evidence from picture naming. *Language and Cognitive Processes*, 22:1114-1139.
- William S. Horton and Richard J. Gerrig (2005a). Conversational Common Ground and Memory Processes in Language Production. *Discourse Processes*, 40(1):1-35.
- William S. Horton and Richard J. Gerrig (2005b). The impact of memory demands on audience design during language production. *Cognition*, 96:127-142.
- William S. Horton and Boaz Keysar (1996). When do speakers take into account common ground? *Cognition*, 59:91-117.
- Falk Huettig and Gerry T. M. Altmann (2005). Word meaning and the control of eye fixation: semantic competitor effects and the visual world paradigm. *Cognition*, 96:B23-B32.
- Claire Hughes and Rosie Esnor (2005). Executive Function and Theory of Mind in 2 Year Olds: A Family Affair? *Developmental Neuropsychology*, 28:645-668.
- Ellen A. Issacs and Herbert H. Clark (1987). References in conversation between experts and novices. *Journal of Experimental Psychology: General*, 116:26-37.
- Boaz Keysar, Dale J. Barr, Jennifer A. Balin, and Timothy S. Paek (1998). Definite reference and mutual knowledge: Process models of common ground in comprehension. *Journal of Memory and Language*, 39:1-20.
- Boaz Keysar, Dale J. Barr, Jennifer A. Balin, and Jason S. Brauner (2000). Taking perspective in conversation: The role of mutual knowledge in comprehension. *Psychological Science*, 11(1):32-38.
- Boaz Keysar, Shuhong Lin, and Dale J. Barr (2003). Limits on theory of mind use in adults. *Cognition*, 89:25-41.
- Boaz Keysar (2007). Communication and miscommunication: The role of egocentric processes. *Intercultural Pragmatics*, 4:71-84.

- Robert M. Krauss and Sidney Weinheimer (1966). Concurrent feedback, confirmation, and the encoding of referents in verbal communication. *Journal of Personality and Social Psychology*, 4(3):343–346.
- Edmundo Kronmüller and Dale J. Barr (2007). Perspective-free pragmatics: Broken precedents and the recovery-from-preemption hypothesis. *Journal of Memory and Language*, 56:436–455.
- Willem J. M. Levelt (1989). *Speaking*. MIT Press, Cambridge, Massachusetts.
- Shuhong Lin, Boaz Keysar, and Nicholas Epley (2010). Reflexively mindblind: Using theory of mind to interpret behavior requires effortful attention. *Journal of Experimental Social Psychology*, 46:551-556.
- Maryellen C. MacDonald and Morten H. Christiansen (2002). Reassessing Working Memory: Comment on Just and Carpenter (1992) and Waters and Caplan (1996). *Psychological Review*, 109(1):35-54.
- James S. Magnuson, James A. Dixon, Michael K. Tanenhaus, and Richard N. Aslin (2007). The dynamics of lexical competition during spoken word recognition. *Cognitive Science*, 31:133-156.
- Danielle Matthews, Elena Lieven, and Michael Tomasello (2010). What's in a manner of speaking? Children's sensitivity to partner-specific referential precedents. *Developmental Psychology*, 46:749-760.
- Ken McRae, Michael J. Spivey-Knowlton, and Michael K. Tanenhaus (1998). *Journal of Memory and Language*, 38:283-312.
- Charles Metzing and Susan E. Brennan (2003). When conceptual pacts are broken: Partner-specific effects on the comprehension of referring expressions. *Journal of Memory and Language*, 49:201-213.
- Antje S. Meyer, Astrid M. Sleiderink, and Willem J. M. Levelt (1998). Viewing and naming objects: eye movements during noun phrase production. *Cognition*, 66:B25-B33.
- Aparna S. Nadig and Julie C. Sedivy (2002). Evidence of perspective-taking constraints in children's on-line reference resolution. *Psychological Science*, 13(4):329-336.
- Mante S. Nieuwland, Marte Otten, and Jos J. A. Van Berkum (2007). Who are you talking about? Tracking discourse-level referential processing with event-related potentials. *Journal of Cognitive Neuroscience*, 19(2):228-236.
- Elizabeth S. Nilsen and Susan A. Graham (2009). The relations between children's communicative perspective-taking and executive functioning. *Cognitive Psychology*, 58:220-249.
- Jared M. Novick, Sharon L. Thompson-Schill, and John C. Trueswell (2008). Putting lexical constraints in context into the visual-world paradigm. *Cognition*, 107:850-903.
- Thomas Pechmann, T. (1989). Incremental speech production and referential overspecification. *Linguistics*, 27:89–110.
- Daniel C. Richardson, Rick Dale, and Natasha Z. Kirkham (2007). The art of conversation is coordination. *Psychological Science*, 18(5):407-413.
- Craige Roberts (2003). Uniqueness in definite noun phrases. *Linguistics and Philosophy*, 26:287-350.
- Rachael D. Rubin, Sarah Brown-Schmidt, Melissa C. Duff, Daniel Tranel, and Neal J. Cohen (2009). Common ground representations in hippocampal amnesia. Poster presented at the Society for Neuroscience conference, Chicago, IL.
- Susanne Salmon-Alt and Laurent Romary (2000). Generating Referring Expressions in Multimodal Contexts. Paper presented at the Workshop on Coherence in Generated Multimedia - INLG 2000, Mitzpe Ramon, Israel.
- Michael F. Schober (1993). Spatial perspective-taking in conversation. *Cognition*, 47:1-24.
- Michael F. Schober (1995). Speakers, Addressees, and Frames of Reference: Whose Effort is

- Minimized in Conversations about Locations? *Discourse Processes*, 20:219-247.
- Michael F. Schober and Susan E. Brennan (2003). Processes of Interactive Spoken Discourse: The role of the partner. *Handbook of discourse processes*, eds. A. C. Graesser, M. A. Gernsbacher, & S. R. Goldman, pages 123-164. Lawrence Erlbaum, Hillsdale, New Jersey.
- Michael F. Schober and Herbert H. Clark (1989). Understanding by addressees and overhearers. *Cognitive Psychology*, 21:211-232.
- Michael F. Schober and Frederick G. Conrad (2008). Survey interviews and new communication technologies. In Frederick G. Conrad and Michael F. Schober (Eds.), *Envisioning the survey interview of the future*, pages 1-30. Wiley, New York, New York.
- Rose Scott, Sarah Brown-Schmidt, Cynthia Fisher, and Renee Baillargeon (2009). 2.5-year-olds Consider Others' False Beliefs When Interpreting Their Speech. Paper presented at the Society for research on child development annual conference, Denver, CO.
- Julie C. Sedivy, Michael K. Tanenhaus, Craig G. Chambers, and Greg N. Carlson (1999). Achieving incremental semantic interpretation through contextual representation. *Cognition*, 71:109-147.
- Jesse Snedeker and Sylvia Yuan (2008). Effects of prosodic and lexical constraints on parsing in young children (and adults). *Journal of Memory and Language*, 58:574-608.
- Michael J. Spivey and Michael K. Tanenhaus (1998). Syntactic Ambiguity Resolution in Discourse: Modeling the Effects of Referential Context and Lexical Frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24:1521-1543.
- Michael J. Spivey, Michael K. Tanenhaus, Kathleen M. Eberhard, and Julie C. Sedivy (2002). Eye movements and spoken language comprehension: Effects of visual context on syntactic ambiguity resolution. *Cognitive Psychology*, 45:447-481.
- Michael K. Tanenhaus, Michael J. Spivey-Knowlton, Kathleen M. Eberhard, and Julie C. Sedivy (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, 268:1632-1634.
- Michael K. Tanenhaus, Michael J. Spivey-Knowlton, and Joy E. Hanna (2000). Modeling thematic and discourse context effects on syntactic ambiguity resolution within a multiple constraints framework: Implications for the architecture of the language processing system. In Matthew W. Crocker, Martin Pickering, and Charles Clifton, Jr. (Eds.). *Architecture and mechanisms of the language processing system*. Cambridge University Press, Cambridge, United Kingdom.
- Michael K. Tanenhaus, James S. Magnuson, Delphine Dahan, Craig Chambers (2000). Eye Movements and Lexical Access in Spoken-Language Comprehension: Evaluating a Linking Hypothesis between Fixations and Linguistic Processing. *Journal of Psycholinguistic Research*, 29:557-580.
- Michael K. Tanenhaus (2004). On-line sentence processing: past, present and, future. *The on-line study of sentence comprehension: ERPs, eye movements and beyond*, eds. Manuel Carreiras and Charles Clifton, Jr., pages 371-392. Psychology Press, New York, New York.
- Michael K. Tanenhaus, Austin F. Frank, T. Florian Jaeger, Anne Pier Salverda, Mikhail Masharov (2008, March). The Art of the State: Mixed Effects Regression Modeling in the Visual World. Paper presented at the 21st Annual CUNY Conference on Human Sentence Processing, Chapel Hill, North Carolina.
- Steve Tauroza and Desmond Allison (1990). Speech Rates in British English. *Applied Linguistics*, 11:90-105.
- John C. Trueswell and Michael K. Tanenhaus (1994). Toward a lexicalist framework for constraint-based syntactic ambiguity resolution. *Perspectives on Sentence Processing*, eds. Charles Clifton, Jr., Lyn Frazier, and Keith Rayner, pages 155-179. Lawrence Erlbaum Associates, Hillsdale, New Jersey.

- John C. Trueswell, Irena Sekerina, Nicole M. Hill, and Marian L. Logrip (1999). The kindergarten-path effect: Studying on-line sentence processing in young children. *Cognition*, 73:89-134.
- Jos J. A. Van Berkum (2009). Language and affect: how comprehension depends on how we feel and what we care about. 15th Annual Conference on Architectures and Mechanisms for Language Processing (AMLaP 2009). Barcelona, Spain.
- Duane G. Watson, Michael K. Tanenhaus, Christine Gunlogson (2008). Interpreting pitch accents in on-line comprehension: H* vs L H*. *Cognitive Science*, 32:1232-1244.
- Justine B. Wells, Morten H. Christiansen, David S. Race, Daniel J. Acheson, Maryellen C. MacDonald (2009). Experience and sentence processing: Statistical learning and relative clause comprehension. *Cognitive Psychology*, 58:250-271.
- Deanna Wilkes-Gibbs and Herbert H. Clark (1992). Coordinating beliefs in conversation. *Journal of Memory and Language*, 31:183-194.
- Michael P. Wilson and Susan M. Garnsey (2009). Making simple sentences hard: Verb bias effects in simple direct object sentences. *Journal of Memory and Language*, 60:368-392.
- Shali Wu and Boaz Keysar (2007). The effect of culture on perspective taking. *Psychological Science*, 18(7):600-606.
- Eiling Yee and Julie C. Sedivy (2006). Eye movements reveal transient semantic activation during spoken word recognition. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 32:1-14.