Visual world studies of conversational perspective taking: Similar findings, diverging interpretations

Dale J. Barr
University of Glasgow

> What we have proposed is that when a listener tries to understand what a speaker means, the process he goes through can limit memory access to information that is common ground between the speaker and his addressees. [...] ...the comprehension process must keep track of common ground, and its performance will be optimal if it limits its access to that common ground. Whether its design is actually optimal in this respect is a question that can only be answered empirically. (Clark & Carlson, 1981, p. 76-77)

With this *restricted access* hypothesis, Clark and Carlson (1981) set the theoretical stage for what is now an active area of research on perspective taking in spoken language comprehension. Along with similar notions such as mutual knowledge and common knowledge (e.g., Lewis, 1969), common ground—information that interlocutors share and believe they share (Clark & Marshall, 1981)—has played an critical role in theories of pragmatics. Such constructs appear most prominently in Clark's collaborative model, but are characteristic in general of pragmatic approaches that invoke assumptions of cooperativity, assumptions that can be traced back to Grice (1957).

Before the visual world paradigm became the dominant methodology for studying conversational perspective taking, there were a number of investigations using traditional psycholinguistic methodologies (for review, see Barr & Keysar, 2006). A number of these studies found evidence for the use of common ground during language interpretation (Clark, Schreuder, & Buttrick, 1983; Gerrig & Littman, 1990; Greene, Gerrig, McKoon, & Ratcliff, 1994; Gibbs, Mueller, & Cox, 1988), but their methodologies offered only limited insight. First, some of them used third-party judgments of the interpretation of written text (Gerrig & Littman, 1990; Gibbs et al., 1988), which are unlikely to be representative of what takes place when addressees comprehend spoken language in conversational contexts. Second, some of them failed to distinguish the use of information because it was jointly available from its use because of its availability to the self, possibly leading to the underestimation of egocentric language processing (see Keysar, 1997 for discussion). But the critical limitation of these studies is that they offered almost no insight into the time-course with which listeners integrate information about a speaker's perspective with the incoming speech.

Visual-world eyetracking greatly expanded the potential for insight into how listeners access and use common ground during situated language comprehension (for background on visual-world eyetracking, see Spivey and Huette, this volume; Pyykkönnen and Crocker, this volume). Indeed, it is arguably in the study of situated language understanding that the key advantages of the visual world paradigm are most fully realized. First, visual world tasks fundamentally involve reference, with listeners following instructions to manipulate objects or pictures in a display. Referential ambiguity has long been a primary focus of pragmatic theories (Clark & Carlson, 1981; Clark & Marshall, 1981), and is one of the most common sources of misunderstanding in conversation (Schegloff, 1987). Also, the use of spoken language to search for referents within an environment can be quite naturally made part of a joint task, such as working together to rearrange objects in a grid (Hanna, Tanenhaus, & Trueswell, 2003; Nadig & Sedivy, 2002; Keysar, Barr, Balin, & Brauner, 2000). Next, visual world eyetracking makes it possible to observe the referential process as it unfolds in time, without the observation process itself influencing comprehension, since it requires no deliberative judgments beyond those involved in the interpretation process itself. Finally, the listener's gaze location is sampled at such a high rate (60–2000 Hz) as to provide a nearly continuous profile of the entire interpretation process, from anticipatory processing to high-level decision processes. However, with this new observational power comes substantive interpretive and statistical challenges that are often underappreciated.

The study of conversational perspective taking in spoken language comprehension has become a productive area of visual world research, with close to 30 published visual world studies to date in just over a decade of research. From the earliest few studies with typically-developing adults, the area has expanded to investigate perspective taking in a broad range of contexts, including:

- development (Epley, Morewedge, & Keysar, 2004; Nadig & Sedivy, 2002; Sobel, Sedivy, Buchanan, & Hennessy, 2011);
- scalar implicature (Grodner & Sedivy, 2011; Heller, Grodner, & Tanenhaus, 2008);
- the role of executive control (Brown-Schmidt, 2009b; Lin, Keysar, & Epley, 2010);
- effects of bilingualism (Rubio-Fernández & Glucksberg, 2011);
- effects of mood (Converse, Lin, Keysar, & Epley, 2008);
- effects of familiarity between interlocutors (Savitsky, Keysar, Epley, Carter, & Swanson, 2011);
- joint action (Hanna & Tanenhaus, 2004);
- cross-cultural differences (Wu & Keysar, 2007);
- autism (Begeer, Malle, Nieuwland, & Keysar, 2010);
- disfluency (Arnold, Hudson Kam, & Tanenhaus, 2007);
- repeated reference and "conceptual pacts" (Barr & Keysar, 2002; Brennan & Hanna, 2009; Brown-Schmidt, 2009a; Horton & Slaten, 2011; Kronmüller & Barr, 2007; Metzing & Brennan, 2003)

My goal in the current chapter is not to review these studies; many of them have already been reviewed elsewhere (Barr & Keysar, 2006; Brennan & Hanna, 2009; Brown-Schmidt & Hanna, 2011). Instead, my aim is to address some long-standing controversies in this research area that, in my view, have impeded progress on important theoretical issues, and on which there is a pressing need to arrive at some kind of consensus. Recently, several researchers (Brennan & Hanna, 2009; Brown-Schmidt & Hanna, 2011; see also Brown-Schmidt, this volume) have taken note of the apparent divergence of findings in the area, with some studies appearing to show stronger and earlier effects of common ground than others. They explain these divergent findings in terms of the different paradigms various researchers have used to study conversation and the extent to which these paradigms provide interlocutors with opportunities to interactively establish common ground. They further suggest that studies showing the earliest and strongest effects of common ground are those that allow common ground to be established through live interaction. In contrast, they suggest that studies lacking live interaction—studies which often use pre-recorded materials with elaborate cover stories to convince listeners they are listening to live speech—tend to show weaker effects of common ground.

In this chapter, I will argue against such attempt to reconcile findings based on assumptions about interactivity. My argument has two main strands. First, although it seems plausible that interaction gives stronger evidence for common ground, thus potentially yielding stronger effects, the studies targeted as insufficiently interactive do indeed show reliable effects of common ground; however, they do not show such effects on all levels of processing. So any explanation that invokes interactivity would have to say why it is the case that this information was used at some but not at all levels. Second—and more importantly—the explanation accepts the divergence in published findings at face value. However, a closer look at several key studies shows that the divergences are more likely to reflect inconsistent practices of analysis and interpretation applied to underlying body of data that is, in fact, surprisingly consistent. Had all datasets been analyzed in the same way, researchers would have largely come to the same conclusions. Until disagreements about the relationship between theory and data are resolved, it is premature to debate wider issues about the pros and cons of different research paradigms for investigating common ground. If researchers not in agreement about what effects are present in their data, and what such effects mean, attempts to debate any broader issues will be largely futile.

To a large extent, how we analyze data is informed by our theoretical outlook—we look for those things we expect to find, in the manner we expect to find them. To date, research on perspective taking has focused on explanations pitched at the level of individual language users, asking if speakers or listeners use common ground in their processing of language. To demonstrate such person-level effects, it is sufficient to show that common ground had an effect on behavior (or brain activity). However, I will argue that the appropriate level of explanation is not the level of the individual person, but the individual *process*. Thus, rather than asking questions like, *Are listeners sensitive to context in understanding references?* we should be asking questions like, *Does context influence lexical, semantic, syntactic, and/or phonological processing; and if so, how?*

This focus on process-level explanations also calls for a different approach in how we analyze and interpret data from visual-world eyetracking studies. To adequately support claims about effects of context on particular processes requires experimentally isolating those processes in the data. I will present evidence that many of the diverging findings in the field are the result of different approaches to the handling of *anticipatory baseline effects* (ABEs) in the analysis of visual world data. ABEs arise in perspective-taking studies using visual occlusion because listeners have access to constraining information about what speakers do and do not know well before they hear referential expressions. Listeners can make use of this constraining information to reduce the set of referential alternatives before hearing speech. To be sure, this undeniably shows that listeners are sensitive to common ground from the earliest moments of processing. However, it is an independent question whether this information is also available to individual processes within the comprehension system—it is entirely possible for individual processes within a cognitive architecture to be unaffected by information actively represented at other levels within the system (Fodor, 1983; Sloman, 1996). Thus, access does not imply *integration*, since it is possible that the information about the speaker's knowledge is not used later to modulate the processing of incoming speech.

It is possible to distinguish between access and integration using appropriate statistical or experimental controls. Fortunately, most studies in the area include the relevant experimental controls, although such controls are sometimes not treated appropriately in the analysis. By looking at the data in a way that includes such controls, I will show that several key studies show similar temporal profiles of common ground use during the interpretive process: early anticipatory effects, followed by bottom-up effects of lexical processing that are not modulated by common ground, followed (optionally) by late effects of common ground that may be post-lexical in nature. Furthermore, this temporal profile for common ground radically differs from the profile of contextual effects induced by verb semantics. Together, these findings are consistent with the proposal that lexical processes are encapsulated from common ground (and possibly from other situational sources of constraint), but cannot be straightforwardly accounted for by probabilistic constraint-based approaches.

Visual world studies have quickly become a primary source of data not only in the study of reference resolution and perspective taking, but also in many other areas of spoken language processing. The controversies that have emerged within this particular area of language processing are symptomatic of the more general absence of clear standards for relating visual world data to psycholinguistic theory. Thus, the lessons that can be drawn by considering this area of research in depth are relevant to visual world researchers at large.

The key tests of perspective-taking in spoken language comprehension have come in the form of visual world studies using joint referential communication tasks, in which potential referents are made to be either privileged (known only to the listener) or shared (known both to the listener and speaker) by visual occlusion or by manipulating listeners' beliefs about the speaker. In this review, I will focus on studies using typically-developing adult populations, and that involve the interpretation of simple referential descriptions. After a brief review of the area, I will attempt to reconcile the findings by scrutinizing data from three studies that have similar experimental designs but that differ in social interactivity: Barr (2008b), Hanna et al. (2003), and Brown-Schmidt (2009b). Studies involving repeated reference and "conceptual pacts" (Brennan & Clark, 1996) are not considered here, as they involve additional theoretical and interpretive issues relating to priming and memory (see Kronmüller and Barr, 2015, for a meta-analysis and review).

Theoretical and Empirical Background

In recent discussions of perspective taking in spoken language comprehension, the restricted access hypothesis of Clark and Carlson (1981) is not seen as a serious contender. I believe this reflects the progress in the area, rather than any intrinsic implausibility of the hypothesis in itself. As we have seen from the quote with which we began this chapter, Clark and Carlson made principled theoretical claims that a language processor that limited itself to information in common ground would be maximally efficient. To be sure, the scope of processing that would be restricted in such a way was left vague; it is unclear whether it is intended to apply all the way down to low-level phonological and lexical processing. But it is beyond dispute that at the very least, Clark and colleagues intended the analysis to apply to high-level interpretive processes, such as those involved in interpreting reference: "Demonstrative reference is perhaps the prototype of expressions that cannot be understood without appeal to context. But what context? If our proposal is correct, all the information the listener should ever appeal to is the speaker's and addressee's common ground." (Clark et al., 1983, p. 99). Importantly, it is with respect to these referential processes that the restricted access model has been repeatedly disconfirmed.

An early study by Keysar et al. (2000) used a task in which listeners sat facing a (confederate) speaker and followed his spoken instructions to rearrange objects in a grid that stood vertically between them. Some of the slots of the grid were open from both sides so that their contents could be mutually viewed (making the contents shared), while others were closed off from the speaker's side so that the listener, but not the speaker, could see the contents (making the contents privileged). Some of the spoken instructions mentioned a shared "target" object in a way that also matched a privileged "competitor" object. For example, for one item the listener saw three candles of increasing size, the larger two of which were shared, and the smallest of which was privileged. According to restricted search, when listeners were told to "put the small candle next to the toothpaste," they should only consider the smaller of the two candles visible to the speaker, and not the privileged candle, because the speaker was ignorant of the latter candle's existence. Disconfirming this prediction, listeners attended far more to a privileged small candle than to a privileged toy monkey, and showed severe delays in identifying and selecting the target. In fact, listeners erroneously selected the privileged object instead of the target about 20% of the time, a rate that is surprisingly high, at least from the standpoint of restricted search. Other studies using a computerized version of the task in which listeners do not receive feedback observe an even higher rate of errors, around 40–50% of trials containing a competitor (Apperly et al., 2010).

Later studies sought a more stringent test of restricted search by making it even less plausible that the speaker might know about the contents of occluded squares, and by having the privileged competitor visually inaccessible to the listener (Keysar, Lin, & Barr, 2003). In one experiment, listeners were presented with a grid containing only shared objects and given a box with objects that they were supposed to "hide" from the speaker in the privileged squares. In this way, there could be little doubt that the speaker was unaware of the contents of the occluded spaces. Furthermore, one of these objects (the privileged competitor) was additionally to be placed inside of a bag so that it was no longer visible to the listener, such that looking at it could not be interpreted in terms of low level visual interference. The basic findings were replicated, with longer gazes on privileged competitors than to noncompetitors, severe delays in identifying and selecting the target, and a high error rate (again, around 20%). A second experiment went even further by comparing the standard condition in which the speaker was presumed to be ignorant of the contents of occluded squares to one in which listeners were led to believe that the speaker had a *false belief* about the contents of the square containing the privileged competitor (or noncompetitor); for example, believing that it was a toy truck (noncompetitor) when it was actually a small candle. In spite of the fact that this should have increased the salience of common ground, there was no evidence that listeners were less egocentric in this condition than in the standard ignorance condition.

These studies, in addition to further studies using the same paradigm (Epley et al., 2004; Lin et al., 2010) were presented as evidence against restricted search and in support of an anchoring-and-adjustment model of perspective taking known as *perspective adjustment*. According to perspective adjustment, comprehension processes are initially “anchored” in information available to the self. Listeners can optionally use common ground to adjust away from this anchor point, but this adjustment step is optional, and requires sufficient time and processing resources. Thus, comprehension will be egocentrically biased to the extent that listeners fail to adjust away from their own perspective.

While these initial studies provided clear evidence against restricted search, and documented an alarming degree of egocentrism in spoken language comprehension, they had a number of limitations that subsequent studies sought to address. First, Keysar and colleagues provided only minimal time-course information, such as first and final fixation times. These are only crude measures of online processing, and may not be as sensitive as analyses that test for effects across various time windows.

A further criticism was that the competitors in privileged ground were always a better match to the semantics of the target description than the target itself (Nadig & Sedivy, 2002; Hanna et al., 2003); for instance, the privileged small candle was even smaller than the target small candle. The original rationale for this feature was that it provided a distinct interpretation in privileged ground; otherwise, if the privileged competitor was an equally good fit to the description as the target, then listeners would be forced to use common ground to resolve the ambiguity. It could be argued, however, that this feature leads to an overestimation of listener egocentrism. Nevertheless, even when the target and competitor are equalized for their fit to the referring expression, egocentric behavior is still observed: in one such study, the presence of a competitor caused 65% of listeners to ask for clarification (*which candle?*) at least once during the experiment, even though there was only one possible referent in common ground (this result did not hold for Asian participants; see Wu & Keysar, 2007 and Wu, Barr, Gann, & Keysar, 2013 for additional discussion).

A more serious criticism was that these early efforts did not provide definitive support for the perspective-adjustment view, because they lacked a critical control (Hanna et al., 2003; Nadig & Sedivy, 2002). The analyses always compared a privileged competitor to a privileged noncompetitor. While such a comparison is sufficient to test the restricted search hypothesis, it is insufficient to support perspective adjustment as an alternative. Perspective adjustment assumes that listeners are “egocentric first”, but Keysar, Barr and colleagues only demonstrated that privileged competitors were fixated more that privileged noncompetitors. Showing that listeners were initially egocentric would have required demonstrating that privileged competitors were fixated just as much as competitors in common ground, but the studies lacked this condition. Thus, data from these studies are consistent not only with “egocentric first” models, but also with models which assume that common ground exerts an immediate but partial (rather than absolute) effect on referential processing.

Partial, immediate effects of common ground could be explained by probabilistic constraint-based models (PCBMs). The PCBM approach is thoroughly interactive and nonmodular, and assumes that the different sources of constraint available to the comprehension system, including common ground, is weighted and interactively combined from the earliest moments of comprehension (MacDonald, Pearlmutter, & Seidenberg, 1994; Tanenhaus, Spivey-Knowlton, & Hanna, 2000). Importantly, there is assumed to be no limitation on the interaction between different levels of processing: information at very high levels of processing (such as the systems tracking mutual knowledge) can, in principle, constrain the operation of the lowest levels of processing (e.g., phonological processing and lexical access). The extent of this constraint depends not on the type of information but only on how heavily it is weighted (i.e., its salience and reliability). In this respect, PCBMs are similar in spirit (and often functionally equivalent) to Bayesian models, which mathematically specify the optimal combination of information in probabilistic reasoning (Jurafsky, 1996).

PCBMs assume gradient effects of common ground, and thus predict that less competition should be observed from a competitor in privileged ground than in common ground. Such gradient effects would falsify the “egocentric first” prediction of perspective adjustment. To test this, Hanna et al. (2003), varied whether the critical alternative was privileged or shared. In the study, pairs were visually separated by a divider, and a (confederate) director instructed a listener to place geometric shapes in an array to match the pattern viewed by the director. Instead of visual occlusion, common ground was established through a grounding process in which the director and participant talked about which shapes they had in common. At some point, the director gave a critical instruction describing a target *red triangle*, in the context of a critical alternative that was either a competitor (another red triangle) or noncompetitor (a green triangle), and that was either privileged or shared. Consistent with PCBMs, listeners were more likely to gaze at a shared target than at a privileged competitor, and this difference was present from the earliest moments of comprehension. Similar findings were reported by Nadig and Sedivy (2002) in a study involving five- and six- year-old children. Taken together, these findings disconfirm the “egocentric first” prediction of the perspective-adjustment view.

Brown-Schmidt (2009b) found additional evidence for early effects of common ground. Listeners answered a speaker’s questions about privileged objects (see also Brown-Schmidt, Gunlogson, & Tanenhaus, 2008). The questions included ambiguous nouns that referenced certain shared landmark objects, adjacent to which these privileged objects were located. The ambiguous nouns were disambiguated by a following subordinate phrase: for example, listeners might hear *What’s above the cow that’s wearing shoes?* in a context with two cartoon cows, a “target” landmark cow wearing shoes and a “competitor” landmark cow wearing glasses. The question was whether listeners could, prior to the disambiguating word (e.g., shoes), use common ground to identify the target landmark and associated privileged target. There were two critical manipulations, the first of which, “mention”, was whether the speaker had already sought information about the identity of the privileged competitor located adjacent to the competitor landmark (e.g., the cow wearing glasses), or had instead asked about a control object. In the former case (the “competitor-mentioned” condition), when speakers later asked “What’s above the cow...” listeners could use common ground to infer that the speaker must be asking about the target cow, since she already knew what was above the competitor cow. Brown-Schmidt also introduced a second manipulation, “grounding”, crossed with the first, which was whether or not the speaker gave evidence of actually having properly understood the listener’s reply. If listeners use common ground, they should show the earliest disambiguation effect when a competitor was mentioned and successfully grounded, since this is the case where the evidence was strongest that the speaker already knew the identity of the privileged item that was next to the competitor landmark.

In her analysis, Brown-Schmidt considered three consecutive 400 ms bins starting from the onset of the noun (e.g., *cow*), the first two of which would capture pre-disambiguation effects, and the third of which would capture post-disambiguation effects. The analysis suggested that listeners gazed at the target landmark and adjacent privileged target earlier when the competitor landmark had been mentioned, and did so prior to the disambiguating word. However, this effect only reached significance when the grounding of the privileged competitor had been successful, suggesting that listeners used common ground to resolve the reference.

Reconciling the findings

The above selective review of key studies on perspective taking in spoken language comprehension reveals clear progress, but the field is far from reaching agreement on the nature or timing of partner-specific effects. The main points of agreement are that (1) comprehension is not restricted to common ground, but shows egocentric effects and (2) common ground can be accessed early, and not just as part of a post-comprehension stage, as suggested by the perspective-adjustment model. These findings can be explained by PCBMs. However, although these studies have shown early *access* to common ground, they have not gone further to show that this information was actually *integrated* with subsequent referential processing.

Visual world studies of information integration seek to understand how contextual evidence modulates the uptake of linguistic evidence. Each trial in a visual world study has a particular temporal structure whose importance is often overlooked: namely, that the presentation of the relevant contextual evidence temporally precedes that of the critical linguistic evidence, often by a large interval. For example, in the classic study of effects of visual context on syntactic processing by Tanenhaus, Spivey-Knowlton, Eberhard, and Sedivy (1995), listeners had visual access to the information in the scene for some substantial amount of time prior to hearing the critical, syntactically ambiguous portion of the expression. Or, in perspective taking studies, listeners are given evidence about which referents are shared and which are privileged long before they hear a speaker make reference to any target object. Of course, such a time lag is necessary, given that it is only possible to test whether people make use of potentially constraining contextual information if they are given sufficient time to attend to and process that information. However, prior availability of context can also produce differences in gaze probabilities before the onset of the referring expression, and such *anticipatory baseline effects* (ABEs) can cloud the interpretation of effects present during the acoustic lifetime of the referring expression (Barr, 2008a, 2008b; Barr, Gann, & Pierce, 2010).

ABEs are especially likely to arise in studies of perspective taking, since the critical contextual information about what is or is not in common ground (or who is going to speak next) is nearly always available from the onset of the trial. In one of the first experiments, Keysar et al. (2000) noted that in a five second window prior to speech onset, listeners were more likely to gaze at shared than at privileged referents. Such a "head start" for the probability of gazing at shared objects may persist into the critical referring expression itself. What is important about this is that it reflects information that listeners access *in anticipation* of upcoming speech. At the person level, observing that listeners are more likely to gaze at objects consistent with context could be taken as evidence they are making predictions about what the speaker might refer to next. In this sense, it supports the idea of early sensitivity to common ground—but only at the person level.

Pitching explanations at the process level rather than at the person level opens up the possibility of dissociations—some levels of processing may have access to information that is not accessible at other levels (Fodor, 1983; Sloman, 1996). Indeed, the presence of dissociations between access and integration can be quite informative about underlying cognitive and neurological architecture. Such access-integration dissociations are often seen in vision, for example; knowing that the two lines in the Müller-Lyer illusion are actually of equal length (e.g., by measuring them using a ruler) does not keep us from experiencing them as if they are different; although the equality is represented in our minds, our visual system behaves as though it lacks access to it, a phenomenon known as *cognitive impenetrability* (Fodor, 1983). To show that information was integrated at a particular level of processing, it is insufficient to show that it was attended to at the person level.

To test claims about whether contextual information is *integrated* into linguistic processing at a particular level, it is necessary to statistically or experimental isolate effects at that level in order to assess whether they are modulated by contextual information (Barr, 2008b). Showing that listeners are more likely to look at shared competitors than at privileged competitors—as shown by Hanna et al. (2003) and Nadig and Sedivy (2002), among others—indicates *that* common ground was used, but it doesn't tell you *how* it was used. It is entirely possible that listeners used common ground to *anticipate* what the speaker would refer to next, but were unable to *integrate* that information during certain levels of referential processing. To the extent that gazes to common ground objects are no higher during referential processing as before that processing began casts doubt on the idea that common ground is actually being used in the processing of the expression.

To test these ideas, Barr (2008b) conducted three perspective-taking experiments using a design similar to Hanna, Trueswell, and Tanenhaus, but examining temporary lexical ambiguities (buckle-bucket) rather than full lexical ambiguities. Because the ambiguity is temporary, listeners can ultimately resolve the ambiguity based on the phonology itself. Therefore, finding evidence that common ground modulates the processing of the initial portion of the word in this case would be strong evidence that it is accessed and used spontaneously during spoken language comprehension. Listeners viewed computerized displays containing four objects and heard a speaker (presumed to be speaking from another room and looking at a different computer screen) instruct them to *Click on the [target object]*. In addition to the target object (e.g., bucket), two of the remaining objects on the screen were also in common ground with the speaker. The fourth, final object in each test display was a *critical alternative* that was either a competitor (e.g., buckle) or noncompetitor (e.g., stepladder) and was furthermore either privileged (i.e., the listener believed that the speaker saw a blank box where the listener saw the critical alternative) or shared (i.e., the listener believed that the speaker also could see the object).

The analysis was time-aligned to the onset of the noun identifying the target object (e.g. "bucket"). To the extent that listeners attend to common ground prior to the onset of the noun, during this same interval they should show a tendency to gaze at the critical alternative more when it is shared than when it is privileged. If listeners are able to further *integrate* this information into language processing, then the effect of lexical competition (e.g., whether the critical alternative is a *buckle* or a *stepladder*) should matter more when the critical alternative is in common ground than when it is privileged. In other words, the competition effect (the greater tendency to gaze at the buckle than the stepladder) should be larger when the critical alternative is in common ground, a pattern we will call *anticipation plus integration.* In contrast, if lexical processes are encapsulated from this higher level information, then the competition effect should appear no different in the two conditions, a pattern we will call *anticipation without integration*.

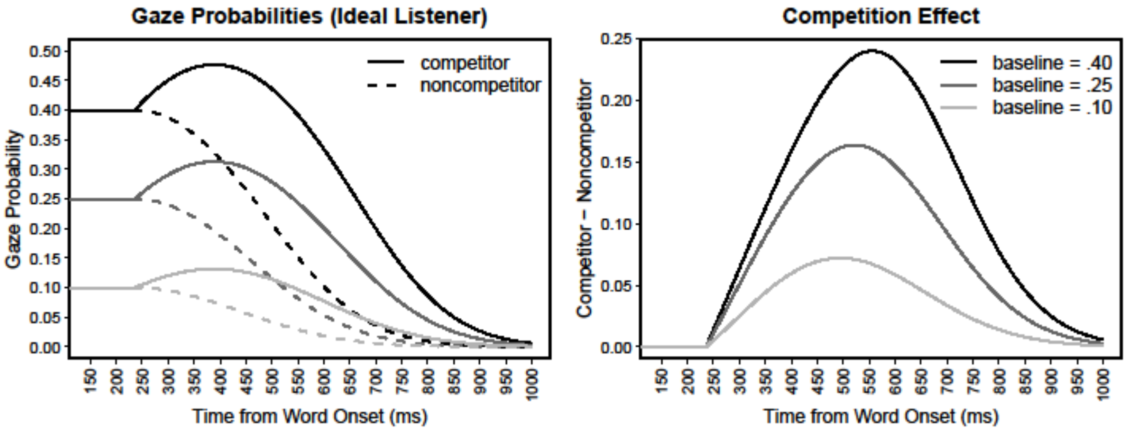

Figure 1. Predicted gaze behavior (left panel) and lexical competition effects (right panel) from an "ideal listener" model under different levels of contextual constraint.

What is the basis for the claim that integration of common ground should lead to attenuation of the lexical competition effect? To sharpen intuitions, let us consider language processing from the point of view of an *ideal listener* who optimally integrates prior expectations with incoming linguistic information. Using Bayes' theorem, we can formalize our intuitions about how prior expectations might modulate the lexical competition effect. Note that lexical competition is defined here as the difference in probability of gazing at the critical alternative when it is a lexical competitor (e.g., *buckle*) versus a noncompetitor (e.g., *stepladder*) when processing the target word (e.g., *bucket*). Figure 1 presents hypothetical data for a fixed evidence function under three different levels of prior expectations (.40, .25, and .10). All of the functions were derived by applying Bayes' theorem to the same evidence under different priors. When the listener believes that the critical alternative is a highly plausible referent—for instance, with a prior probability of .40—there is a large competition effect (right panel). As the critical alternative becomes less plausible, the competition effect becomes smaller; compare the effects for .10 and .40 in the right panel. This is a consequence of the evidence function being multiplied by different priors, as Bayes' Theorem stipulates. (Note that for the purpose of this article, we are dealing with claims about the size of the competition effect on the proportional scale, not the log odds scale.)

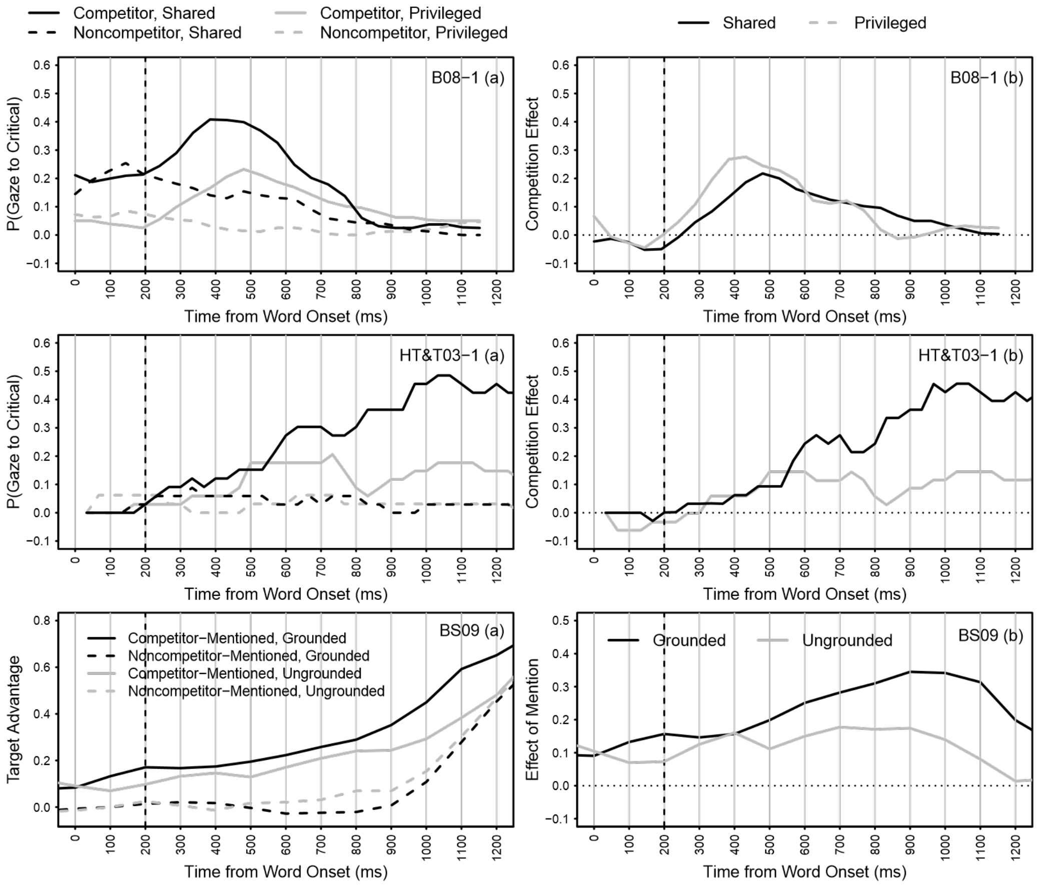

Figure 2. Overall results from Barr (2008), Experiment 1 (B08-1); Hanna, Tanenhaus, & Trueswell (2003), Experiment 1 (HT&T03-1), and Brown-Schmidt (2009) (BS09). Panel (a) shows the results broken down by condition; panel (b) shows the competition effect by common ground status.

In Barr's first experiment (Figure 2, top panel) common ground had a strong, statistically reliable effect on listeners' prior expectation about what the speaker would refer to. Listeners gazed more at the critical alternative when it was in common ground with the listener than when it was privileged (with prior probabilities of approximately .20 versus .05, respectively, from 0–200 ms after speech onset, which is before language driven effects can appear assuming a 200 ms overhead for saccadic programming). But despite a strong difference in prior expectation, there was little evidence for any attenuation whatsoever of the competition effect in the privileged condition (top row, right panel of the figure). This result, which suggests that lexical processes are cognitively impenetrable with respect to common ground, was replicated in two additional experiments.

The second experiment contrasted listeners' ability to integrate common ground with their ability to integrate information from a preceding verb. Based on previous results from Dahan and Tanenhaus (2004), it was expected that the verb-based constraint would induce an anticipation-plus-integration pattern, reflecting the penetrability of lexical processes to prior linguistic context, while the constraint from common ground would yield an anticipation-without-integration pattern, suggesting cognitive impenetrability. One group of participants were in the common ground condition, and completed an experiment identical to Experiment 1. For a second group, the common ground manipulation was replaced with manipulation of the verb preceding the critical noun. Half of the utterances given to this latter group began with the verb *click* (e.g., "click on the bucket"), which is unconstraining inasmuch as it could potentially apply to any picture in the display. This unconstraining-verb condition played a role analogous to the shared condition for the common ground group (in fact, it was identical to that condition). In the other half of the sentences presented to this second group, the verb *click* was replaced by a constraining verb that accepted the target as a potential direct object, but not the critical alternative. For example, the verb *empty* in *empty the bucket* could apply only to the bucket, but not to the stepladder or buckle. This constraining-verb condition plays a role analogous to the privileged ground condition in Experiment 1, because as in that condition, well before the onset of the noun, the contextual information already favors the target over the critical alternative.

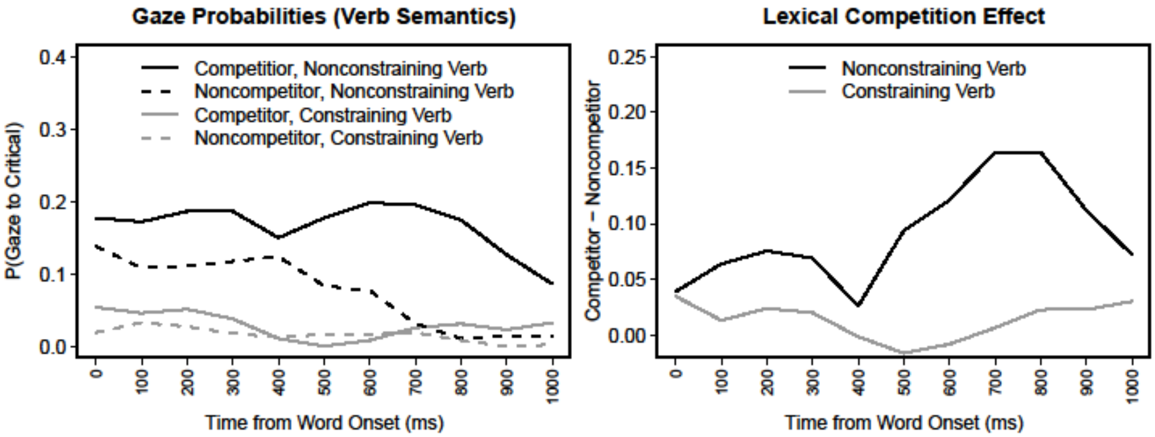

Figure 3. Effects of verb constraint on lexical competition, Experiment 2 of Barr (2008).

The listeners in the common ground condition showed the same anticipation-without-integration pattern seen in the first experiment. In contrast, listeners in the verb-based constraint group showed a very different pattern. Like the common ground group, the prior contextual constraint exerted a strong effect on the prior likelihood of gazing at the critical alternative. When the verb was one for which the critical alternative would be implausible as a direct object (e.g., the verb *empty* in relation to a buckle or a stepladder), listeners had a much lower prior expectation that the critical alternative would be mentioned (about .05 versus .15 in the condition where the verb was not constraining; see Figure 3, left panel). This immediate constraining effect of verbs is similar to that observed in Altmann & Kamide, 1999 and Dahan & Tanenhaus, 2004. The fact that this strong anticipatory effect was present suggests that the verb had been processed in time to influence the processing of the following referential expression. But unlike for the common ground condition, there was strong evidence for attenuated lexical competition in the condition with the lower prior (Figure 3, right panel). In other words, in line with the ideal listener model, whether or not the critical alternative was a competitor mattered less when the critical alternative was an implausible object of the verb.

The anticipatory effects of common ground, which were seen across three independent experiments, supports the claims of Hanna et al. (2003) and Nadig and Sedivy (2002) that listeners are sensitive to common ground from the earliest moments of processing, and thus also reject the “egocentric first” account put forward by Keysar, Barr and colleagues. However, looking at the data from a process level suggests that the egocentrism that is observed across all of these studies may not reflect partial use of common ground, but may be the result of the failure to integrate common ground into lexical processing, despite a strong prior expectation. This might imply that lexical processes are encapsulated from common ground information, and perhaps other kinds of situational information as well. The results from the verb-constraint condition provide an important contrast, inasmuch as they show that lexical processes are not fully encapsulated from all contextual information. Indeed, verb semantics have very strongly constraining effects on processing. It is an important question for future research to characterize the source of differences between these two types of contextual constraint.

Advocates of PCBMs have argued against these findings on both theoretical and statistical grounds. On the theoretical side, Brown-Schmidt and Hanna (2011) cite the lack of interactivity in Barr’s experiments. To be sure, although listeners were actually hearing recorded materials, two of three of the experiments employed elaborate cover stories to convince listeners that they were listening to speakers who spoke to them live from another room. However, Brown-Schmidt and Hanna (2011) believe this is insufficient for establishing common ground:

> ...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." (p. 22).

Note that some of the interactive experiments cited by Brown-Schmidt and Hanna created *opportunities* for grounding, but did not require it as part of the procedure (Nadig & Sedivy, 2002; Heller et al., 2008), and it seems unlikely that referents were actually grounded in every case. This also ignores the fact that grounding is a procedure for dealing with cases where *common ground is uncertain*, and Barr's experiments used procedures where the common ground was made clear through the structure of the "game" participants were playing. When common ground is not uncertain, it is unnecessary (and in fact, odd) to first engage in grounding (e.g., akin to asking someone sitting at your dinner table, "Do you see the salt?" prior to asking them to pass it to you.)

It is also important to note that listeners in Barr's experiments clearly *attended* to common ground: indeed, the odds of gazing at common ground referents before the onset of the expression were three to four times higher than the odds of gazing at privileged ground referents. Because the interactional affordances account assumes that grounding is necessary to form common ground, it could best account for Barr's results if there were *null effects of common ground altogether*, but it is not clear how it would explain why effects of common ground were very large for some processes (referential anticipation) but negligible for others (lexical competition).

The second criticism leveled against Barr's results is statistical in nature, and calls in question the use of regression to partial out effects of common ground on anticipation versus integration. This criticism first appeared in a conference presentation by Tanenhaus and colleagues (Tanenhaus, Frank, Jaeger, Masharov, & Salverda, 2008) and was further discussed by Brown-Schmidt and Hanna (2011) in a footnote. The approach in Barr (2008a) was to parameterize a polynomial regression model so that baseline effects were captured in the intercept term, with the time-varying (e.g., slope) parameters capturing integration effects. Tanenhaus and colleagues acknowledged that anticipatory baseline effects pose a threat to interpretation, but suggested that Barr's approach may introduce statistical artifacts. They note that the gaze-state a listener is in at the onset of the referring expression constrains possibilities for subsequent gaze states, with a particular concern about cases where at the start of the referring expression, listeners are already looking at the referential alternative being analyzed. It is difficult to go into detail about the nature of these arguments as they have not been adequately expounded in the literature, with only a one page conference abstract publically available. Given the scanty information, it is not clear at all whether gaze-state dependencies are claimed to exist as an analytical fact or as an empirical possibility. A passing remark in Brown-Schmidt (2009b) suggests that she considers it an analytical fact having to do with the nature of eye data, particularly due to the fact that "...one cannot make a saccade to what one is already looking at" (p. 896). The relevance of this tautological observation is unclear since the primary events being analyzed are *eye gazes*, which extend over time, not discrete saccadic events. It is indeed possible to continue or discontinue gazing at what one is already gazing at (see Barr, Gann, & Pierce, 2011 for further discussion). Such "in principle" analytical concerns about gaze-state dependences seem ill-founded. Still, it is also reasonable to view gaze-state dependencies as an empirical possibility, in which the sequences of gazes leading up to the referring expression somehow influences how the referential expression itself is processed. Frank, Salverda, Jaeger, and Tanenhaus (2009) reported evidence for such dependencies, but their results may have been artifacts due to regression toward the mean; ruling out this possiblity would require a fuller evaluation of their methods than is possible from a single conference poster presentation. In short, it is premature to dismiss the statistical solution proposed by Barr (2008a) until: (1) convincing empirical or theoretical arguments in favor of the gaze state dependencies have been made in the scientific literature; (2) the logic of how such dependencies might bias the proposed statistical solution is clearly articulated and validated; and (3) the magnitude of any claimed biases have been measured and compared to the potential biases of any alternative solutions. At the time of writing, none of these conditions has been met.

One thing that has been ignored in the debate over baseline effects is the fact that it is not necessary to control for baseline effects statistically; indeed, they can be controlled experimentally by using appropriate control conditions. Indeed, such experimental control was already built into the design of Barr (2008b). The analysis did not directly compare gaze on a shared buckle to gaze on a privileged buckle, but instead compared the effect of competition (competitor vs. noncompetitor) *within* the privileged condition (privileged buckle versus privileged stepladder) to competition *within* the shared condition (shared buckle versus shared stepladder). Note that *within* the privileged condition, there is no baseline effect: the privileged noncompetitor starts off at the same probability as the privileged competitor. By the same logic, there is no baseline difference between the shared competitor and shared noncompetitor. Furthermore, it cannot be argued that because the shared competitor starts higher than the privileged competitor, it has less room to go up, artificially dampening the competition effect for shared referents.[1] This apparent "disadvantage" for the shared competitor is *perfectly offset* by the "advantage" that the shared noncompetitor has *more room to go down* (thus potentially enhancing the competition effect) than the privileged noncompetitor.

The top right panel of Figure 2 (labeled B08-1b) compares these competition effects (competitor minus noncompetitor for each of the shared and privileged conditions), controlling for anticipatory baseline differences. This analysis very clearly shows no evidence whatsoever for a larger competition effect in the common ground condition (if anything, the effect appears larger in the privileged condition). Given that this evidence exploits experimental rather than statistical control, arguments about "gaze-state dependencies" do not apply.

Finally, the invocation of interactive grounding as an explanation for the different findings is questionable because there may be no differences there to explain in the first place, given that not all available studies distinguish access from integration. Still, it is possible to qualitatively assess whether the findings are concordant by considering probability data

[1] Barr et al. (2011) evaluated the logic of this argument about conditions with higher baselines having less room to go up and found it lacking. They showed that regardless of whether trials are on- or off-region at the onset of referring expression, they have equal potential, in principle, to increase or decrease subsequent target probabilities. Off-region trials vote in favor of the target by becoming on region, and against it by staying off; on-region trials vote in favor of the target by staying on region, and against it by becoming off.

presented in the figures of Hanna et al. (2003) and Brown-Schmidt (2009b). The probability data was extracted and reconstructed from the figures in these papers by a pixel mapping technique using photo editing software (GIMP).

Hanna et al. (2003) used an identical 2x2 factorial design to Barr (2008b), in which competition (competitor vs. noncompetitor) was crossed with common ground status (privileged vs. shared). Although this design allows for the experimental control of anticipatory effects, the authors' analyses did not take advantage of this control. Instead, inferential statistics were presented for data from just one cell of the design, where the critical alternative was a privileged competitor. This analysis directly compared probability of gazing at the target (which was in common ground) to the probability of gazing at the privileged competitor in this same condition, over a 200-800 ms window. However, the advantage they found for the shared target could be explained entirely as the result of anticipatory baseline effect favoring the target. If we consider data from the full design, would we see a smaller competition effect in the privileged ground condition, as the ideal listener model predicts? And if so, at what point would the effect appear?

Gaze probabilities to the critical alternative from Experiment 1 of Hanna et al. (2003) were extracted from Figures 2 and 3 of their manuscript, and are given in the middle row of Figure 2 of the current manuscript. Analyzing the data in the same manner as we just did for Barr's experiment, we find an anticipation-without-integration pattern (middle row, right panel). Clearly, there is an overall competition effect starting at around 250-300 ms, as evident in the rise of the lines from zero in both the shared and privileged condition. However, the competition effect in the privileged condition seems *identical* to the shared condition until around 600 ms at which it begins to diverge. There is a simple reason why this divergence appears in Hanna et al. but not in Barr's experiments: Barr used lexical ambiguities that were quickly resolved by the input (buckle-bucket), whereas Hanna et al. used full ambiguities (e.g., both target and competitor were identical red triangles). Because the input never resolved the ambiguity, listeners in the shared competitor condition would have to ask the speaker which referent she intended; thus, the competition effect in this condition is long lasting. In contrast, in the privileged competitor condition, listeners could spontaneously resolve the ambiguity by making use of the information that the speaker was unaware of the privileged red triangle. The fact that they did this late—well after the onset of the competition effect—suggests that the effect could be postlexical, and is thus consistent with the evidence presented in Barr (2008b) for the encapsulation of lexical processing from common ground.

A similar approach can be applied to results from Brown-Schmidt (2009b). Following the approach described by Tanenhaus et al. (2008), Brown-Schmidt (2009b) removed all trials starting with a gaze to either the target or competitor regions (nearly 40% of all data). However, Barr et al. (2011) showed that not only is such drastic data removal unnecessary, it actually introduces bias due to regression toward the mean (and potentially selection biases as well). Therefore, we consider the results for the full data that Brown-Schmidt presented in the appendix (Figure A1 of Brown-Schmidt, 2009b). Unlike the previous analyses, which looked at competition effects, here we look at the effect of "mention", that is, of whether or not the listener had already attempted to give the speaker information about the identity of the privileged item adjacent to the competitor landmark. Note that it is only in the "grounded" condition that listeners should consider this information part of common ground, because it is only in this condition that listeners had evidence that speakers actually registered the information.

The raw data are plotted in the left panel of the bottom row of Figure 2, with the effect of mention plotted in the right panel of the same row. One notable result is that even before listeners processed the target word (e.g., *cow*), there was already a quite substantial effect of mention present in both the grounded and ungrounded condition. In fact, at 0 ms (the onset of the word) the effect of mention already looks slightly larger in the grounded condition than in the ungrounded condition. Thus, even before listener knew that the speaker would refer to the cow, they were already paying attention to information in common ground. This apparent anticipatory baseline effect is entirely consistent with that observed in Barr's experiments. Note additionally that the difference between the effects of mention for the grounded and ungrounded conditions only really begins to exceed this baseline effect 600-800 ms after word onset; the lines seem to rise roughly in parallel up to this point. This overall pattern—an apparent anticipatory effect of common ground, followed by apparent-partner independent processing, followed again by a late effect of common ground—is consistent with encapsulated language processing during the ambiguous noun.

Different interpretations of the same underlying data pattern

In summary, literature reviews of visual world studies on perspective taking have largely taken the diverging findings of the various studies at face value, and some have attributed these apparently different findings to differences in the extent to which the paradigms used by different labs afford collaborative interaction (Brennan & Hanna, 2009; Brown-Schmidt & Hanna, 2011; see also Brown-Schmidt, this volume). Such authors suggest that those studies in which common ground is established interactively are also those that show the strongest effects of common ground. But this explanation seems implausible, for two reasons. First, it leaves unexplained why some noncollaborative studies show strong effects of common ground on certain aspects of processing (i.e., anticipatory baseline effects) but not on others (i.e., competition effects). Second, and more importantly, they make the mistake of assuming that the divergent findings are real. However, when anticipatory effects of common ground are controlled for, these studies show roughly the same evidence in favor of the encapsulation of lexical processes from common ground, *regardless of the collaborative potential afforded by the paradigm*.

This analysis reveals that the failure to appropriately distinguish access from integration has led to the overestimation of listeners' abilities to integrate common ground with incoming input. Whereas listeners seem to be able to integrate semantic information from a preceding verb to a near optimal level, this does not seem to be the case for common ground. The results generally suggest that there is a period early in the processing of referential expressions that proceeds entirely autonomously from common ground, and possibly from other kinds of situational information. But given the controversial nature of this claim, it is important to pursue further studies corroborating the basic finding, as well as attempting to delimit the types of contextual information that cannot be integrated.

The study of perspective taking in language processing is challenging on many levels. Researchers often adopt conflicting definitions of what counts as perspective taking or common ground, sometime conflating notions of mutual belief with shared information or information that is merely associated with a speaker (see Keysar, 1997 and Lee, 2001 for discussion). Theoretical disputes arise out of a failure to distinguish the use of *speaker associated* information from the use of meta-representational information about a speaker's beliefs. For instance, an ERP study has shown that stereotypical information associated with a particular type of speaker influences lexical processing (van Berkum, van den Brink, Tesink, Kos, & Hagoort, 2008); e.g., listeners experience a classic N400 effect to the contextually inappropriate word "tattoo" when hearing the sentence "I have a large tattoo on my back" spoken in an upper-class accent. But stereotypical information about a type of speaker is not the same as information about a particular speaker's beliefs and goals; the former type of information is representational and contextually stable; the latter is meta-representational and can be highly contextually variable. Additionally, studying perspective taking or "mentalizing" more generally is challenging because many things that look like genuine mentalizing can be produced by simpler mechanisms that do not involve representations of another's beliefs (for discussion, see Barr, 2014; Heyes, 2014). It is also a problem that interlocutor behavior in highly interactive contexts is mutually dependent (by definition), which makes it difficult to distinguish behaviors that reflect mutual adjustments arising from feedback from truly individual cognitive adaptations undertaken unilaterally and spontaneously (Barr, 2014).

Finally, as noted in this review, research on perspective taking is challenging because of the often complex nature of the relationship between data and theory, which arises from the rich nature of visual-world eyetracking data. Despite this complexity, the fact that researchers are asking increasingly sophisticated and nuanced questions about perspective taking is an encouraging sign of progress. However, to progress further, the field needs to forge consensus on basic issues of data analysis and interpretation. The approaches that researchers adopt to data analysis in visual-world perspective-taking studies are currently far too eclectic, unprincipled, and ad hoc. Unfortunately, this is probably also true of visual world research in general. Statistical and experimental solutions have already been proposed in the peer-reviewed literature (Barr, 2008a,b; Barr et al., 2011), but researchers often ignore these solutions without adequate justification based on the suspicion that they are unsound. However, this suspicion currently lacks a clear theoretical or empirical justification. Citing vague concerns about possible "gaze state dependencies" should not give researchers *carte blanche* to ignore the interpretive problems imposed by anticipatory baseline effects, nor to dismiss the solutions to these problems that have already been proposed and evaluated. Resolving this debate should be prioritized, as a basic consensus on analysis and interpretation is preliminary to any broader theoretical debates about interactivity and language processing. Research in this area still has great potential to enhance our understanding of language processing in real-world settings, but can only do so if it rests on a solid foundation of data analysis and interpretation.

References

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: restricting the domain of subsequent reference. *Cognition*, *73*, 247–264.

Apperly, I. A., Carroll, D. J., Samson, D., Humphreys, G. W., Qureshi, A., & Moffitt, G. (2010). Why are there limits on theory of mind use? evidence from adults' ability to follow instructions from an ignorant speaker. *The Quarterly Journal of Experimental Psychology*, *63*, 1201–1217.

Arnold, J. E., Hudson Kam, C. L., & Tanenhaus, M. K. (2007). If you say it 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.

Barr, D. J. (2008a). Analyzing 'visual world' eyetracking data using multilevel logistic regression. *Journal of Memory and Language*, *59*, 457–474.

Barr, D. J. (2008b). Pragmatic expectations and linguistic evidence: Listeners anticipate but do not integrate common ground. *Cognition*, *109*, 18–40.

Barr, D. J. (2014). Perspective-taking and its impostors in language use: Four patterns of deception. In T. Holtgraves (Ed.), *The oxford handbook of language and social psychology* (pp. 98–110). New York: Oxford University Press.

Barr, D. J., Gann, T. M., & Pierce, R. S. (2011). Anticipatory baseline effects and information integration in visual world studies. *Acta Psychologica*, *137*, 201–207.

Barr, D. J., & Keysar, B. (2002). Anchoring comprehension in linguistic precedents. *Journal of Memory and Language*, *46*, 391–418.

Barr, D. J., & Keysar, B. (2006). Perspective taking and the coordination of meaning in language use. In M. J. Traxler & M. A. Gernsbacher (Eds.), *Handbook of psycholinguistics (2nd ed.)* (pp. 901–938). Amsterdam, Netherlands: Elsevier.

Begeer, S., Malle, B. F., Nieuwland, M. S., & Keysar, B. (2010). Using theory of mind to represent and take part in social interactions: Comparing individuals with highfunctioning autism and typically developing controls. *European Journal of Developmental Psychology*, *7*, 104–122.

Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *22*, 1482– 1493.

Brennan, S. E., & Hanna, J. E. (2009). Partner-Specific adaptation in dialog. *Topics in Cognitive Science*, *1*, 274–291.

Brown-Schmidt, S. (2009a). Partner-specific interpretation of maintained referential precedents during interactive dialog. *Journal of Memory and Language*, *61*, 171–190.

Brown-Schmidt, S. (2009b). The role of executive function in perspective taking during online language comprehension. *Psychonomic Bulletin & Review*, *16*, 893–900.

Brown-Schmidt, S., Gunlogson, C., & Tanenhaus, M. K. (2008). Addressees distinguish shared from private information when interpreting questions during conversation. *Cognition*, *107*, 1122-1134.

Brown-Schmidt, S., & Hanna, J. E. (2011). Talking in another's shoes: Incremental perspective-taking in language processing. *Dialogue and Discourse*, *2*, 11–33.

Clark, H. H., & Carlson, T. B. (1981). Context for comprehension. In J. Long & A. Baddeley (Eds.), *Attention and performance ix* (pp. 313–330). Hillsdale, N. J.: Erlbaum.

Clark, H. H., & Marshall, C. R. (1981). Definite reference and mutual knowledge. In A. K. Joshe, B. L. Webber, & I. A. Sag (Eds.), *Elements of discourse understanding* (pp. 10–61). Cambridge: Cambridge University Press.

Clark, H. H., Schreuder, R., & Buttrick, S. (1983). Common ground and the understanding of demonstrative reference. *Journal of Verbal Learning & Verbal Behavior. Vol*, *22*, 245–258.

Converse, B. A., Lin, S., Keysar, B., & Epley, N. (2008). In the mood to get over yourself: Mood affects theory-of-mind use. *Emotion*, *8*, 725–730.

Dahan, D., & Tanenhaus, M. K. (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.

Epley, N., Morewedge, C. K., & Keysar, B. (2004). Perspective taking in children and adults: Equivalent egocentrism but differential correction. *Journal of Experimental Social Psychology*, *40*, 760–768.

Fodor, J. A. (1983). *The modularity of mind: An essay on faculty psychology*. Cambridge, MA: MIT Press.

Frank, A. F., Salverda, A. P., Jaeger, T. F., & Tanenhaus, M. K. (2009). Multinomial models with “state” dependencies. In *CUNY 2009 Conference on Human Sentence Processing.*

Gerrig, R., & Littman, M. (1990). Disambiguation by community membership. *Memory & Cognition*, *18*, 331–338.

Gibbs, R. W., Mueller, R. A. G., & Cox, R. W. (1988). Common ground in asking and understanding questions. *Language and Speech*, *31*, 321–335.

Greene, S., Gerrig, R., McKoon, G., & Ratcliff, R. (1994). Unheralded pronouns and management by common ground. *Journal of Memory and Language*, *33*, 511–511.

Grice, H. P. (1957). Meaning. *The philosophical review*, *66*, 377–388.

Grodner, D., & Sedivy, J. C. (2011). The effect of speaker-specific information on pragmatic inferences. In E. A. Gibson & N. J. Perlmutter (Eds.), *The processing and acquisition of reference* (pp. 239–271). Cambridge, MA: MIT Press.

Hanna, J. E., & Tanenhaus, M. K. (2004). Pragmatic effects on reference resolution in a collaborative task: Evidence from eye movements. *Cognitive Science*, *28*, 105–115.

Hanna, J. E., Tanenhaus, M. K., & Trueswell, J. C. (2003). The effects of common ground and perspective on domains of referential interpretation. *Journal of Memory and Language*, *49*, 43–61.

Heller, D., Grodner, D., & Tanenhaus, M. K. (2008). The role of perspective in identifying domains of reference. *Cognition*, *108*, 831–836.

Heyes, C. (2014). Submentalizing: I am not really reading your mind. *Perspectives on Psychological Science*, *9*, 131–143.

Horton, W., & Slaten, D. (2011). Anticipating who will say what: The influence of speakerspecific memory associations on reference resolution. *Memory & Cognition*, 1–14.

Jurafsky, D. (1996). A probabilistic model of lexical and syntactic access and disambiguation. *Cognitive Science*, *20*, 137–194.

Keysar, B. (1997). Unconfounding common ground. *Discourse Processes*, *24*, 253–270.

Keysar, B., Barr, D. J., Balin, J. A., & Brauner, J. S. (2000). Taking perspective in conversation: The role of mutual knowledge in comprehension. *Psychological Science*, *11*, 32–38.

Keysar, B., Lin, S., & Barr, D. J. (2003). Limits on theory of mind use in adults. *Cognition*, *89*, 25–41.

Kronmüller, E., & Barr, D. J. (2007). Perspective-free pragmatics: Broken precedents and the recovery-from-preemption hypothesis. *Journal of Memory and Language*, *56*, 436–455.

Kronmüller, E., & Barr, D. J. (2015). Referential precedents in spoken language comprehension: A review and meta-analysis. *Journal of Memory and Language, 83*, 1–19.

Lee, B. P. H. (2001). Mutual knowledge, background knowledge and shared beliefs: Their roles in establishing common ground. *Journal of Pragmatics*, *33*, 21–44.

Lewis, D. (1969). *Convention: A philosophical study*. Cambridge, M.A.: Harvard University Press.

Lin, S., Keysar, B., & Epley, N. (2010). Reflexively mindblind: Using theory of mind to interpret behavior requires effortful attention. *Journal of Experimental Social Psychology*, *46*, 551–556.

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). Lexical nature of syntactic ambiguity resolution. *Psychological Review*, *101*, 676–703.

Metzing, C., & Brennan, S. E. (2003). When conceptual pacts are broken: Partnerspecific effects on the comprehension of referring expressions. *Journal of Memory and Language*, *49*, 201–213.

Nadig, A. S., & Sedivy, J. C. (2002). Evidence of perspective-taking constraints on children's on-line reference resolution. *Psychological Science*, *13*, 329–336.

Rubio-Fernández, P., & Glucksberg, S. (2011). Reasoning about other people's beliefs: Bilinguals have an advantage. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

Savitsky, K., Keysar, B., Epley, N., Carter, T., & Swanson, A. (2011). The closenesscommunication bias: Increased egocentrism among friends versus strangers. *Journal of Experimental Social Psychology*, *47*, 269–273.

Schegloff, E. (1987). Some sources of misunderstanding in talk-in-interaction. *Linguistics*, *25*, 201–218.

Sloman, S. S. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, *119*, 3–22.

Sobel, D. M., Sedivy, J., Buchanan, D. W., & Hennessy, R. (2011). Speaker reliability in preschoolers' inferences about the meanings of novel words. *Journal of Child Language*, *39*, 90–104.

Tanenhaus, M. K., Frank, A., Jaeger, T. F., Masharov, M., & Salverda, A. P. (2008). The art of the state: Mixed-effect regression modeling in the visual world. In *CUNY 2008 Conference on Human Sentence Processing.*

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*, 1632.

Tanenhaus, M. K., Spivey-Knowlton, M. J., & Hanna, J. E. (2000). Modeling thematic and discourse context effects on ambiguity resolution within a multiple constraints framework: Implications for the architecture of the language processing system. In M. W. Crocker, M. Pickering, & Clifton, Jr (Eds.), *Architectures and mechanisms for language processing* (pp. 90–118). Cambridge: Cambridge University Press.

van Berkum, J. J., van den Brink, D., Tesink, C. M., Kos, M., & Hagoort, P. (2008). The Neural Integration of Speaker and Message. *Journal of Cognitive Neuroscience*, *20*, 580–591.

Wu, S., Barr, D. J., Gann, T. M., & Keysar, B. (2013). How culture influences perspective taking: differences in correction, not integration. *Frontiers in Human Neuroscience*, *7*, 822.

Wu, S., & Keysar, B. (2007). The effect of culture on perspective taking. *Psychological Science*, *18*, 600–606.