In J. M. Henderson, and F. Ferreira (Eds.), *The interface of language, vision, and action: Eye movements and the visual world*. New York: Psychology Press.

Referential domains in spoken language comprehension:

Using eye movements to bridge the product and action traditions*

Michael K. Tanenhaus

University of Rochester

Craig G. Chambers

University of Calgary

Joy. E. Hanna

State University of New York at Stony Brook

Send correspondence to:

Michael K. Tanenhaus

Department of Brain and Cognitive Sciences

Meliora Hall

University of Rochester

Rochester, N. Y. 14627

mtan@bcs.rochester.edu

1.0. Introduction

Most psycholinguistic research on language comprehension falls into one of two traditions, each with its roots in seminal work from the 1960s (Clark, 1992; 1996), and each with its own characteristic theoretical concerns and dominant methodologies. The language-as-product tradition has its roots in George Miller's synthesis of the then emerging information processing paradigm and Chomsky's theory of transformational grammar (e.g., Miller, 1962; Miller & Chomsky, 1963). The product tradition emphasizes the individual cognitive processes by which listeners recover linguistic representations—the 'products' of language comprehension. Psycholinguistic research within the product tradition typically examines moment-by-moment processes in real-time language processing, using fine-grained reaction time measures and carefully controlled stimuli.

The language-as-action tradition has its roots in work by the Oxford philosophers of language use, e.g., Austin, (1962), Grice (1957) and Searle (1969), and work on conversational analysis, e.g., Schegloff & Sachs, 1973). The action tradition focuses on how people use language to perform acts in conversation, arguably the most basic form of language use. Psycholinguistic research within the action tradition focuses primarily on investigations of interactive conversation using natural tasks, typically in settings with real-world referents and well-defined behavioral goals.

It is tempting to view these traditions as complementary; research in the product tradition examines the early perceptual and cognitive processes that build linguistic representations, whereas research in the action tradition focuses on subsequent cognitive and social-cognitive processes that build upon these representations. Although there is some truth to this perspective, it masks some fundamental disagreements. This becomes apparent when we focus on context, a concept that has featured prominently in research in both traditions.

Context in the product and action traditions:

In the product tradition, context is typically viewed either as information that enhances or instantiates a context-independent core representation or as a correlated constraint in which information from higher-level representations can, in principle, inform linguistic processing at lower levels of representation. Specific debates about the role of context include whether, when and how: (a) lexical context affects sub-lexical processing; (b) syntactic and semantic context affect lexical processing; and (c) discourse and conversational context affect syntactic processing. Each of these questions involves debates about the architecture of the processing system and the flow of information between different types of representations--classic information processing questions. Because linguistic processing occurs incrementally as a sentence is read or heard, answering these questions requires response measures that are closely time-locked to the unfolding utterance. These on-line tasks track changes in processing difficulty, e.g., monitoring fixations in reading (Rayner & Liversedge this volume), or changes in representation, e.g., lexical priming, as linguistic input is encountered.

Consider, for example, the well-known cross-modal lexical priming paradigm. This paradigm builds upon the classic finding that response times to a target word are facilitated when the target is preceded by a semantically related prime word (Meyer & Schvaneveldt, 1971). A schematic of a widely used variant of this task, introduced by Swinney, Onifer, Prather & Hirshkowitz (1978), is illustrated in Figure 1.

Figure 1 about here

The participant, who is wearing headphones, listens to sentences that have been pre-recorded by the experimenter. A sentence or short sequence of sentences is presented on each trial. At some point in the sentence, a target letter string appears on a computer monitor, allowing for experimenter control over the timing of the probe with respect to the input. The participant's task is to make a forced-choice decision indicating whether the letter string is a word. The pattern of decision latencies is used to assess comprehension processes. For example, when the target word follows *testified*, a verb whose subject, *doctor*, has been separated by a relative clause, lexical decisions to words that are associated to the subject are faster compared to lexical decisions to unrelated target words. Context is manipulated by varying the properties of the linguistic stimuli that precede the probe word.

In the action tradition, context includes the time, place and participant's conversational goals, as well as the collaborative processes that are intrinsic to conversation. A central tenet is that utterances can only be understood relative to these

factors. For example, Clark points out that in the utterance, *Look at the stallion*, the expression, *the stallion*, could refer to a horse in a field, a painting of a horse, or a test tube containing a blood sample taken from a stallion, depending upon the context of the utterance. Thus, researchers within the action focus primarily on interactive conversation using natural tasks, typically with real-world referents and well-defined behavioral goals.

Consider, for example, the referential communication task originally introduced by Krauss and Weinheimer (1966). A schematic of a well-studied variant of this task introduced by Clark and his colleagues (e.g., Clark & Wilkes-Gibbs, 1986) is illustrated in Figure 2.

Figure 2 about here

Two naïve participants, a Matcher and a Director, are separated by a barrier. Each has the same set of shapes arranged in different positions on a numbered grid. The participants' goal is for the Matcher to rearrange the shapes on his grid to match the arrangement on the Director's grid. The resulting conversation can then be analyzed to provide insights into the principles that guide interactive conversation. While this paradigm abstracts away from certain aspects of typical conversation, such as a shared visual environment and face-to-face interaction, it preserves many central characteristics, including unscripted dialogue between participants who have clear behavioral goals.

With two such different views about so fundamental a notion as context, one might ask why, with the exception of an occasional shot fired across the bow (e.g., Clark & Carlson, 1986; Clark, 1997), the action and product traditions have not fully engaged one another. One reason is methodological. The traditional techniques in the psycholinguist's toolkit for studying real-time language processing required using either text or pre-recorded audio stimuli in contextually limited environments. One cannot, for instance, use cross-modal lexical priming to examine real-time language comprehension (or production) in a referential communication task.

Recently, the development of accurate, relatively inexpensive head-mounted eye-tracking systems has made it possible to examine real-time spoken language processing in rich behavioral contexts. The use of eye movements in spoken language comprehension was pioneered by Cooper (1974), who demonstrated that the timing of participants' eye movements to pictures was related to relevant information in a spoken story. More recently, Tanenhaus, Spivey-Knowlton, Eberhard & Sedivy (1995) showed that when participants follow spoken instructions to manipulate objects in a task-relevant "visual world", fixations to task-relevant objects are closely time-locked to the unfolding utterance, providing insights into the resolution of referential and syntactic ambiguities. Subsequent studies have shown that this technique provides a continuous real-time measure of comprehension processes at a temporal grain fine enough to track even the earliest moments of speech perception and lexical access (e.g., McMurray, Tanenhaus & Aslin, 2002; Tanenhaus, Magnuson, Dahan & Chambers, 2000). Moreover, a clear quantitative linking hypothesis between underlying processes and

fixations makes it possible to use eye movements to evaluate predictions made by mechanistic models (Allopenna, Magnuson & Tanenhaus, 1998, Dahan, Magnuson & Tanenhaus, 2001). Thus the eye-movement paradigm meets the strictest methodological criteria of researchers in the product tradition. At the same time, the language relates to salient real-world referents and is relevant to achieving behavioral goals. Moreover, one can naturally extend the paradigm to interactive conversation, using referential communication tasks. Thus the eye movement paradigm can be used with tasks that satisfy the methodological criteria of the action tradition.

Although eye movement paradigms enable researchers to extend real-time measurement to more natural linguistic settings, including interactive conversation, it is prudent to ask whether the considerable effort that is required to do so is worthwhile. We believe the investment is likely to prove fruitful for at least two reasons. First, the use of eye-tracking paradigms with natural tasks will allow researchers from both traditions to investigate phenomena that would otherwise prove intractable. Second, and perhaps most importantly, this research is likely to deepen our understanding of language processing by opening up each tradition to empirical and theoretical challenges from the other tradition. For example, the product-based construct of priming provides an alternative mechanistic explanation for phenomena such as lexical and syntactic entrainment, i.e., the tendency for interlocutors to use the same words and syntactic structures, that does not require appeal to the action-based claim that such processes reflect active cooperation among speakers and addressees. Conversely, the action-based notion of context offers a challenge to the product-based assumption that

there are core linguistic processes, e.g., word recognition and syntactic parsing, supported by quasi-independent processing modules that do not take into account context-specific information.

In the remainder of this chapter we focus on definite reference in order to examine how people update referential domains in real-time comprehension. We consider how actions, intentions, real-world knowledge, and mutual knowledge circumscribe referential domains, and how these domains affect syntactic ambiguity resolution. Thus we examine whether variables central to the action view of context influence the earliest moments of syntactic processing, which comprises the core of sentence processing, according to many within the product tradition. **Section 2.0** presents a brief discussion of the notion of domains for definite reference and presents evidence that listeners dynamically update referential domains taking into account the real-world properties of objects and their affordances with respect to task-relevant actions. **Section 3.0** reviews results demonstrating that syntactic ambiguity resolution takes place with respect to these task-relevant referential domains, contrary to claims made by proponents of modularity. **Section 4.0** presents evidence that addressees can use information about speaker perspective to guide even the earliest moments of reference resolution. **Section 5.0** presents preliminary results from an ongoing investigation of real-time reference resolution and generation using a completely unscripted referential communication task, which further highlight the importance of action-based referential domains. **Section 6.0** concludes with a brief discussion of implications and future directions.

2.0. Referential domains and definite reference

Many linguistic expressions can only be understood with respect to a circumscribed context or referential domain. Consider, definite descriptions, such as *the new eye-tracker* in a conversation overheard one morning in late May in East Lansing:

A: What did you guys do after dinner?

B: We stayed up until 2 AM drinking beer and debating whether the new eye-tracker is better than the old one.

Felicitous use of a definite noun phrase requires reference to, or introduction of, a uniquely identifiable entity. For example, if there were two martini glasses in front of you at the table, your dinner partner could not felicitously ask you to, *Pass the Martini glass*. Instead your partner would have to use the indefinite version?, *a Martini glass*. Yet, B's use of the definite noun phrase *the new eye-tracker* is easily understood in the example discourse, despite the fact that the speaker and addressee were aware of many different eye-trackers. The definite expression is felicitous here because the satisfaction of uniqueness must be qualified by an all-important caveat, namely, <u>with respect to a relevant context.</u> A definite noun phrase can be used with multiple potential referents so long as the relevant domain defines a unique interpretation. For example, at a banquet one could ask the person sitting next to you to *please pass the red wine* even if

there were six bottles of the same red wine on the table, but only one was clearly within reach of the addressee.

A considerable body of research demonstrates that listeners dynamically update referential domains based on expectations driven by linguistic information in the utterance (Altmann & Kamide, 1999; Chambers, Tanenhaus, Eberhard, Filip & Carlson, 2002; Eberhard, Spivey, Sedivy & Tanenhaus, 1995). The question we asked is whether referential domains take into account the affordances of potential referents with respect to the action evoked by the instruction. If so, this would provide strong evidence in support of the action-based claim that language processing takes place with respect to a particular behavioral context.

In Chambers et al. (2002, Experiment 2), participants were presented with six objects in a workspace, as is illustrated in Figure 3a. On critical trials, the objects included a large and a small container, e.g., a large can and a small can. The critical variable manipulated in the workspace was whether the to-be-mentioned object, in figure 3, the cube, could fit into both of the containers, as was the case for the small version of the cube, or could only fit into the larger container, as was the case for the large version of the cube. Critical instructions for the display in Figure 3 are given in:

1. Pick up the cube. Now put it inside a/the can.

Figure 3 about here

We will refer to the can containers as the **goal-objects** and the to-be-moved object as the **theme-object**. Thus the size of the theme-object determined whether one or two of the potential goal-objects were compatible referents. The instructions manipulated whether the Goal was introduced with the definite article, *the*, which presupposes a unique referent, or the indefinite article, *a*, which implies that the addressee can choose from among more than one Goal.

First, consider the predictions for the condition with the small cube. Here we would expect confusion when the definite article was used to introduce the Goal because there is not a unique referent. In contrast, the indefinite article should be felicitous because there is more than one referent. This is precisely what we found. Eye-movement latencies to fixate the goal-object chosen by the participant were slower in the definite condition compared to the indefinite condition. This confirms expectations derived from the standard view of how definite and indefinite articles are interpreted. Now consider predictions for the condition with the large cube, the theme-object that would fit into only one of the goal-objects, i.e., the large can. If the referential domain consists of those objects in the visual world that meet the linguistic description in the utterance, then the pattern of results should be similar to that for the small cube, resulting in a main effect of definiteness and no interaction with the compatibility. If, however, listeners dynamically update referential domains to include only those objects that afford the required action, that is containers that the object in hand would fit into, then only the large cube would be in the relevant referential domain. Therefore, use of a definite description, e.g., *the can* should be felicitous, because there is only one can that

the cube could be put into, whereas an indefinite description, e.g., *a can*, should be confusing. Thus, there should be an interaction between definiteness and compatibility. As Figure 3b shows, this what we found. Eye-movement latencies to the referent for the definite referring expressions were faster when there was only one compatible referent compared to when there were two, whereas the opposite pattern occurred for the indefinite expressions. Moreover, latencies for the one-referent compatible condition were comparable to control trials in which there was only a single object that met the referential description in the instruction, e.g., trials with only a single large can. These results demonstrate that referential domains are dynamically updated to take into account the real-world properties of potential referents with respect to a particular action.

To further evaluate the claim that the participant's intended actions were constraining the referential domain, we conducted a second experiment in which we modified the second instruction to make it a question, as in example:

2. Pick up the cube. Could you put it inside a/the can?

In order to prevent participants from interpreting the question as an indirect request, the participant first answered the question. On about half of the trials when the participant answered "yes", the experimenter then asked the participant to perform the action. Unlike a command, a question does not commit the speaker and the addressee to the assumption that the addressee can, and will, perform an action. Thus, the referential domain should now take into account all the potential referents that satisfy the linguistic description, not just those that would be compatible with possible action

mentioned in the question. If referential domains take into account behavioral goals, then under these conditions, definite expressions should be infelicitous regardless of compatibility, whereas indefinite expressions should always be felicitous,. This is what we found. Time to answer the question was longer for questions with definite compared to indefinite referring expressions. Crucially, definiteness did not interact with compatibility (i.e., size of the theme-object). Moreover, compatibility had no effect on response times for questions with definite articles (for details, see Chambers, 2001).

These results demonstrate that referential domains are dynamically updated using information about available entities, properties of these entities, and their compatibility with the action evoked by an utterance. This notion of referential domain is, of course, compatible with the view of context endorsed by researchers in the action tradition. However, it is important to note that assignment of reference necessarily involves mapping linguistic utterances onto entities in the world, or a conceptual model thereof. A crucial question, then, is whether these contextually-defined referential domains influence core processes in language comprehension that arguably operate without access to contextual information. In order to address this question we examined whether action-based referential domains affect the earliest moments of syntactic ambiguity resolution.

3.0. Referential domains and syntactic ambiguity resolution

As the utterances in (3) unfold over time, the underlined prepositional phrase, *on the towel*, is temporarily ambiguous because it could introduce the Goal as in (3a) or modify the Theme, as in (3b).

3. a. Put the apple <u>on the towel</u>, please.

   b. Put the apple <u>on the towel</u> into the box.

Temporary "attachment' ambiguities like these have long served as a primary empirical test-bed for evaluating models of syntactic processing (Tanenhaus & Trueswell, 1995). Crain and Steedman (1985), also Altmann & Steedman (1988), called attention to the fact that many classic structural ambiguities involve a choice between a syntactic structure in which the ambiguous phrase modifies a definite noun phrase and one in which it is a syntactic complement (argument) of a verb phrase. Under these conditions, the argument analysis is typically preferred. For instance, in example (3a), readers and listeners will initially misinterpret the prepositional phrase, *on the towel*, as the Goal argument of *put* rather than as an adjunct modifying the noun phrase, *the apple*, resulting in a garden-path.

Crain and Steedman (1985) noted that one use of modification is to differentiate an intended referent from other alternatives. For example, it would be odd for (3a) to be uttered in a context in which there was only one perceptually salient apple, such as the scene in Figure 5, Panel A, whereas it would be natural in contexts with more than one apple, as in the scene illustrated in Panel B. In this context, the modifying phrase,

on the towel, provides information about which of the apples is intended. Crain & Steedman proposed that listeners might initially prefer the modification analysis to the argument analysis in situations that provided the appropriate referential context. Moreover, they suggested that referential fit to the context, rather than syntactic complexity, was the primary factor controlling syntactic preferences (also cf, Altmann & Steedman).

Tanenhaus et al. (1995) and Spivey, Tanenhaus, Eberhard & Sedivy (2002) investigated the processing of temporarily ambiguous sentences such as (4a) and unambiguous control sentences, such as (4b), in contexts such as the ones illustrated in Figure 4.

Figure 4 about here

The objects illustrated in the figures were placed on a table in front of the participant. Participants' eye movements were monitored as they performed the action in the spoken instruction:

4. a. Put the apple on the towel in the box.

   b. Put the apple that's on the towel in the box.

The results, which are presented in Figure 5, provided clear evidence for immediate use of the visual context. In the one–referent context, participants frequently looked at? the false (competitor) goal, indicating that they initially misinterpreted the prepositional

phrase, *on the towel,* as introducing the Goal. In contrast to the results for the one-referent context, looks to the competitor goal were dramatically reduced in the two-referent context. Crucially, participants were no more likely to look at the competitor goal with ambiguous instructions compared to the unambiguous baseline. Moreover, the timing of the fixations provided showed that the prepositional phrase was immediately interpreted as modifying the object noun phrase (for details see Spivey et al., 2002). Participants typically looked at one of the potential referents as they heard the beginning of the instruction, e.g., *put the apple*. On trials in which participants looked first at the incorrect Theme (e.g., the apple on the napkin), they immediately shifted to the correct Theme (the apple on the towel) as they heard *towel*. The timing was identical for the ambiguous and unambiguous instructions. For similar results, see Trueswell et al., (1999; Trueswell & Gleitman, this volume).

Clearly, then, referential context can modulate syntactic preferences from the earliest moments of syntactic ambiguity resolution. We are now in a position to ask whether the relevant referential domain is defined by the salient entities that meet the referential description provided by the utterance or whether it is dynamically updated based on real-world constraints, including action-based affordances of objects. In Chambers, Tanenhaus & Magnuson (2003), we addressed this issue using temporarily ambiguous instructions such as, *Pour the egg in the bowl over the flour*, and unambiguous instructions, such as, *Pour the egg that's in the bowl over the flour*, with displays such as the one illustrated in Figure 5.

Figure 5 about here

The display for the test trials included the goal (the flour) a competitor goal (the bowl), the referent (the egg in the bowl), and a competitor referent (the egg in the glass). The referent was always compatible with the action evoked by the instruction, e.g., the egg in the bowl was liquid and therefore could be poured. The critical manipulation was whether the affordances of the competitor referent were also compatible with the action evoked by the verb in the instruction. For example, one can pour a liquid egg, but not a solid egg. In the compatible competitor condition, the other potential referent, the egg in the glass, was also in liquid form. In the incompatible competitor condition, it was an egg in a shell. The crucial result was the time spent looking at the competitor Goal, which is presented in Figure 5.

When both potential referents matched the verb (e.g., the condition with two liquid eggs, as in Panel A), there were few looks to the false goal (e.g., the bowl) and no differences between the ambiguous and unambiguous instructions. Thus, the prepositional phrase was correctly interpreted as a modifier, replicating the pattern observed by Spivey et al. (also see Tanenhaus et al., 1995; Trueswell et al., 1999; Trueswell & Gleitman, this volume). However, when the properties of only one of the potential referents matched the verb, (e.g., the condition where there was a liquid egg and a solid egg), we see the same data pattern as Spivey et al. (2002) found with one-referent contexts. Participants were more likely to look to the competitor goal (the bowl) with the ambiguous instruction than with the unambiguous instruction. Thus,

listeners misinterpreted the ambiguous prepositional phrase as introducing a Goal only when a single potential referent (the liquid egg) was compatible with a pouring action.

While these results are problematic for many classes of processing models, it can be argued that they do not provide definitive evidence that non-linguistic constraints can affect syntactic processing. Note that, unlike the Chambers et al. (2002) study, which used the verb *put,* all of the relevant affordances were related to properties that can plausibly be attributed to the semantics of the verb. For example, *pour* requires its Theme to have the appropriate liquidity for pouring. There is precedent going back at least to Chomsky (1965) for incorporating a subset of semantic features, so-called selectional restrictions, into linguistic lexical representations, as long as those features have syntactic or morphological reflexes in at least some languages. Thus it could be argued that only a restricted set of real-world properties influence syntactic processing, viz., those properties that are embedded within linguistic representations stored in the lexicon.

Chambers et al. (2003) addressed this issue in a second experiment. The critical instructions contained *put*, e.g., *Put the whistle (that's) on the folder in the box*, a verb that obligatorily requires a Goal argument. Figure 6 shows a corresponding display, containing two whistle referents, one of which is attached to a loop of string. Importantly, *put* does not constrain which whistle could be used in the action described by the instruction.

Figure 6 about here

The compatibility of the referential competitor was manipulated by varying whether or not participants were provided with an instrument. The experimenter handed the instrument to the participant without naming it. We also avoided semantic associations between the instrument name and the target referent name. For example, before participants were given the instruction described earlier, they might be given a small hook. Critically, this hook could be not be used to pick up the competitor whistle without a string. Thus, upon hearing *put the*..., the competitor could be excluded from the referential domain based on the affordances of the object with respect to the intended action, i.e., using the hook to move an object. If so, participants should misinterpret *on the folder* as the Goal only when ambiguous instructions are used and when an instrument is provided. If, however, the relevant referential domain is defined using only linguistic information, then a Goal misanalysis should occur regardless of whether an instrument is supplied beforehand.

Figure 6 shows the mean time spent fixating the false Goal object within the 2500ms after the first prepositional phrase. The false Goal is most often fixated when ambiguous instructions are used and the competitor cannot afford the evoked action. The remaining conditions all show fewer fixations to the false Goal.

The experiments described in this section provide clear evidence that the syntactic role assigned to a temporarily ambiguous phrase varies according to the number of possible referents that can afford the action evoked by the unfolding instruction. The same results hold regardless of whether the constraints are introduced linguistically by the verb, or non-linguistically by the presence of a task-relevant

instrument. Thus the referential domain for initial syntactic decisions is determined by the listener's consideration of how to execute an action, rather than by information sources that can be isolated within the linguistic system. The results show that provisional syntactic structures are established as part of the process of relating a sentence to the context, taking into account communicative goals. This process requires the simultaneous use of multiple information sources—a property that is consistent with constraint-based theories as well as action-based approaches to language performance. Of critical importance is the nature of the information used in this process. Syntactically-relevant referential context is established by evaluating the affordances of candidate referents against the action evoked by the unfolding sentence. This action itself can be partially determined by situation-specific factors such as the presence of a relevant instrument. The syntactic role assigned to an unfolding phrase in turn depends on whether these factors jointly determine a unique referent without additional information.

These results add to the growing body of literature indicating that multiple constraints can affect even the earliest moments of linguistic processing. Moreover, they present a strong challenge to the claim that the mechanisms underling linguistic processing include encapsulated processing modules. If modular systems are characterized by domain-specificity and isolation from high-level expectations, including perceptual inference (Coltheart, 1999; Fodor, 1983), then the mechanisms underlying on line syntactic processing do not meet these criteria.

4.0. Referential domains and speaker perspective

Thus far we have established that listeners dynamically update referential domains taking into account real world properties of potential referents, including their affordances with respect to intended actions evoked by the unfolding instruction. We now turn to the question of whether a listener's referential domain can also take into account the perspective of the speaker.

Clark and his colleagues have argued that language processing is a form of joint action (Clark, 1996). Interlocutors can cooperate and communicate successfully only if they monitor what is mutually known about a situation and use that knowledge effectively to create **common ground** (e.g., Clark, 1992; 1996; Clark & Brennan, 1989; Clark & Marshall, 1981; Clark & Schaefer, 1987; Clark & Wilkes-Gibbs, 1986). Common ground includes information established on the bases of community membership, physical co-presence, and linguistic co-presence. For example, conversational participants would be able to infer that they share various types of knowledge on the basis of both being in a particular city, or by looking at a particular object at the same time, or by maintaining a record of what has been discussed. A logical consequence of the joint action perspective is that one of the primary roles of common ground is to act as the domain of interpretation for reference.

In real-time comprehension, however, addressees must rapidly access, construct, and coordinate representations from a variety of subsystems. What would constitute optimal use of common ground under these conditions? One possibility is that

conversational partners continuously update their mental representations of common ground, such that a dynamically changing representation of common ground defines the domain within which utterances are typically processed. Updating common ground at this temporal grain would require participants to continuously monitor each other's utterances and actions in order to gather evidence about each other's beliefs. Alternatively, common ground might be interrogated at a coarser temporal grain. If so, many aspects of linguistic processing, including reference resolution, might take place without initial appeal to common ground. Thus, while common ground might control language performance at a macro-level, the moment-by-moment processes necessary for production and comprehension could take place relatively egocentrically for speakers and listeners. Keysar and colleagues' monitoring and adjustment (Horton & Keysar, 1996) and perspective adjustment (Keysar, Barr, Balin, & Brauner, 2000) models adopt this approach.

In the perspective adjustment model (Keysar et al., 2000), the initial interpretation of utterances is egocentric. Common ground is used as a second-stage filter to rule out inappropriate interpretations. A number of theoretical arguments can be marshaled in support of the common ground as filter approach. Computing common ground by building, maintaining, and updating a model of a conversational partner's beliefs could be inefficient because it is extremely memory intensive. In addition, many conversational situations are constrained enough that an individual participant's perspective will provide a sufficient approximation of true common ground. Moreover, information about another's beliefs can be uncertain at best. Thus, adjustment from

monitoring and explicit feedback might be the most efficient mechanism for correcting the occasional confusions or misunderstandings that arise from adopting an egocentric perspective.

Perhaps the strongest evidence that common ground might not initially restrict referential domains comes from a study by Keysar, Barr, Balin, & Brauner (1996, 2000; see also Keysar, Barr, & Horton, 1998). Keysar et al. (2000) monitored eye movements during a referential communication task. Participants were seated on opposite sides of a vertical grid of squares, some of which contained objects. A confederate speaker played the role of director and instructed the naïve participant, who wore a head-mounted eye-tracker, to reorganize the objects in different locations in the grid. Most of the objects were in common ground on the basis of physical co-presence since they were visible from both sides of the display, but a few objects in the grid were hidden from the director's view, and thus were in the matcher's **privileged** ground. On critical trials, the director used a referring expression that referred to a target object in common ground, but that could also refer to a hidden object in privileged ground. For example, one display had three blocks in a vertical row, and the block at the very bottom was hidden. If the director said *Put the bottom block below the apple*, the underlined definite noun phrase referred to the middle block from the common perspective, but to the hidden object from the matcher's egocentric perspective. The results showed that matchers were not only just as likely to look at the hidden object as the target object, but in fact initially preferred to look at the hidden object, and on some trials even picked it up and began to move it.

However, this result does not necessarily support the claim that addressees ignore the speaker's perspective. In the critical conditions, the hidden object was always a better perceptual match for the referring expression than the visible object. For example, when the director said *Put the bottom block...,* the hidden block was the one on the absolute bottom of the display. Similarly, when the director said *Put the small candle …*, the hidden candle was the smallest candle in the display, and the visible candles were medium and large sized.

Hanna, Tanenhaus and Trueswell (in press, Experiment 1) eliminated the typicality confound in a referential communication task using an explicit grounding procedure with common ground established on the basis of linguistic co-presence. Participants wore a head-mounted eye tracker while they played the role of **addressee** in a referential communication task with a **confederate** speaker. For ease of exposition, we use "she" to refer to the addressee and "he" to refer to the confederate. The confederate, who was hidden behind a divider, instructed the naïve participant to manipulate colored shapes on a display board with the pretense of getting the participant's board to match the confederates. The confederate followed a script so that the form of the referring expressions could be controlled and linked to addressees eye movements and actions.

On critical trials, the confederate gave an instruction containing a definite noun phrase that referred to a particular shape as a target location (e.g., *Now put the blue triangle on <u>the red one</u>*). On these trials there were two identical shapes (among others) already in place on the board that could serve as the intended location, the target shape

and the competitor shape (e.g., two triangles). The target shape was always in common ground because it had been referred to and placed on the board during the course of the matching task. We manipulated whether the competitor shape was also in common ground, or whether it was in the addressee's privileged ground. When a shape was in privileged ground, it was identified only to addressee, who placed this "secret shape" in a space on the display board before the matching task began. In addition to manipulating the ground of the competitor shape, the target and competitor shapes were either the same or a different color. A sample display is presented in Figure 7.

Figure 7 about here

The competitor shape was a possible referent (e.g. for *the red one*) only when it was the same color as the target shape. Thus, the different color condition provides a baseline, and looks to the competitor should be infrequent and should not vary as a function of whether the competitor is in common or privileged ground. The critical question was whether a same color competitor would compete when it was in privileged ground as strongly as when it was in common ground. When the same color competitor was in common ground, addressees should consider both the target and competitor shapes as possible referents, and in fact should ask for clarifying information to determine which location was intended by the confederate. The crucial condition was when the same color competitor was in privileged ground. If the language processing system makes immediate use of common ground, addressees should quickly choose the target shape in common ground and there should be little if any interference from the secret shape. However, if common ground is used to filter

initially egocentric interpretations, then addressees should initially consider the secret shape equally as often as the target shape in common ground.

Instructions for each trial were typed on index cards and placed in envelopes. The confederate's instructions contained the configuration of shapes for his board as well as the script he would follow for that trial; the addressee's instructions indicated the secret shape and where to put it on the display board. The addressee's envelope also contained seven shapes, five of which were needed need for the trial. The confederate chose shapes for each trial from a box located on his side of the table.

The confederate's scripts for the critical trials consisted of the following series of instructions: 1) a question asking the addressee what shapes she had in her envelope; 2) a statement telling the addressee which shapes she needed for that trial; 3) three separate statements instructing the addressee to place a shape on the board in a particular location; 4) a critical instruction telling the addressee to stack the final shape on one of the shapes that was already on the board; and 5) a statement that the confederate was finished or, twenty-five percent of the time, a question asking the addressee to describe where her shapes were to double-check that the boards matched. On trials where the target and competitor shapes were both in common ground and were the same color, the addressee typically asked which location was meant. The confederate then apologized for his mistake and told the addressee the specific location.

The results for the different color competitors were as predicted: few looks to the competitor and no effects of type of ground. Figure 8 presents the proportion of fixations on the shapes in the display for the same color competitors in the common and

privileged ground conditions over the course of 2000 ms in 33 ms intervals, beginning with the point of disambiguation, which we defined as the onset of the disambiguating word. At the point of disambiguation, participants were most often fixating on the stacking shape which was either in the resource area or, if they had picked it up already, in their hand. The average offset of the disambiguating region is indicated on each graph.

Figure 8 about here

When there was a same color competitor in common ground (Figure 8a), participants initially looked roughly equally at any of the objects on the board, and within 600 ms after the onset of the disambiguating word began looking primarily at either the target or the competitor shape. The critical question was whether a same color competitor in privileged ground would compete to the same extent as a competitor in common ground. As Figure 8b shows, participants most often made initial looks to the target shape. Within 400 ms after the onset of the adjective, participants were fixating on the target more often than the privileged ground competitor. The proportion of fixations to the target then rose steadily and quickly. The proportion of fixations to the competitor began to rise about 400 ms after the onset of the adjective, and rose along with the proportion of looks to the target until about 800 ms after the onset of the adjective. Note, however, that the proportion of looks to the privileged ground competitor was always lower than the proportion of looks to the target.

The pattern and timing of results demonstrate that (1) addressees used common ground from the earliest moments of reference resolution, and (2) there was some

degree of interference from a potential referent in privileged ground. Our evidence that common ground modulated the earliest moments of reference resolution is consistent with similar studies by Arnold, Trueswell, & Lawentmann (1999) and Nadig & Sedivy (2002). Arnold et al. (1999) conducted an experiment similar to ours with a simplified display that was designed for children. They found that attentive adults experienced competition from a common ground competitor but not from a privileged ground competitor. Nadig and Sedivy (2002) used a display of four physically co-present objects, one of which was hidden from a confederate speaker. Children from 5 to 6 years of age experienced interference from a competitor in common ground but no interference from a privileged ground competitor.

Although the research strategy of comparing the effects of information in common ground and privileged ground has been quite fruitful, this class of experimental design is subject to a potentially problematic ambiguity of interpretation. Privileged ground objects are, by necessity, not referred to by confederates. Therefore, it is unclear whether addressees' preference for referents in common ground arises because they are taking into account the perspective of the speaker, or whether they are performing a kind of probability matching, keeping track of the likelihood with which specific referents are mentioned. Therefore, Hanna et al. (in press) conducted a second experiment to determine the time course with which an addressee can utilize information taken from a speaker's perspective when (1) it conflicts with perceptually salient information in the addressee's privileged ground, and (2) conflicting perspectives are not confounded with likelihood or recency of reference.

Hanna et al. manipulated the perspectives of the conversational participants in a referential communication task such that the domain of interpretation was different from each conversant's perspective; that is, from the speaker's point of view only one subset of objects would make sense as the domain in which reference would be interpreted, while for the addressee a different set of objects would potentially constitute the domain of interpretation. The participant addressee was again seated across a table from a confederate speaker who was completely hidden from view by a vertical divider. (In this case, the confederate speaker was female and will be referred to as she, and participant addressees will be referred to as he.) The experimenter placed four objects on a shelf on the addresee'**s** side of the table, and then named them in order from left to right to inform the confederate of their identity. The confederate repeated the object names in the same order to ground them and firmly establish her perspective, and then instructed the addressee to pick one of them up and place it in one of two areas on the table surface.

The point in the spoken instructions where the referent of the noun phrase became unambiguous was manipulated with respect to the visual display. The point of disambiguation was varied by taking advantage of the different uniqueness conditions carried by definite and indefinite descriptions, and the contrastive property of scalar adjectives (Sedivy, Tanenhaus, Chambers & Carlson, 1999). To give an example using a definite description, consider a display such as the top array of objects in Figure 9 in which there are two sets of two identical objects: two jars, one with olives and one without, and two martini glasses, one with olives and one without. As the instruction

Pick up the empty martini glass unfolds, the addressee cannot identify the referent as the martini glass without olives until relatively late in the instruction. *The empty* could refer to either the empty jar or the empty martini glass in this display, and it is not until the addressee hears *martini* that the referent is disambiguated. Now consider the same instruction in combination with a display such as the middle array of objects in Figure 9 in which both jars and one martini glass are empty, but the other martini glass has olives in it. In this case, the intended referent could be disambiguated upon hearing *the empty*. The definite article *the* signals a uniquely identifiable referent, which eliminates the set of jars as potential referents and locates the referent within the set of martini glasses, and *empty* uniquely identifies the martini glass without olives. Indefinite instructions such as *Pick up one of the empty martini glasses* were also combined with displays to produce late and early points of disambiguation, but we will not discuss the results with indefinites here (for details, see Hanna et al., in press).

Figure 9 about here

The point of disambiguation was either late or early as described above (e.g., *martini* versus *the empty*). We will consider the point of disambiguation to be the onset of the disambiguating word. The perspectives of the participant addressee and the confederate speaker either matched or mismatched. In the matching perspective conditions, the experimenter described the objects to the confederate accurately. In the mismatching perspective conditions, the experimenter described the early disambiguation displays inaccurately (from the addressee's perspective) to the confederate. Mismatching perspectives for both the definite and indefinite instructions

were achieved by describing the objects with the modification switched between the sets. For example, for the definite instruction *Pick up the empty martini glass,* the objects were described as two empty jars, an empty martini glass, and a martini glass with olives in it, but the group really consisted of an empty jar, a jar with olives in it, and two empty martini glasses. See the bottom array of objects in Figures 9 for the mismatching perspective conditions.

With matching perspectives, we expected to see a clear point of disambiguation effect depending on the uniqueness properties of the objects in the display. In the late disambiguation conditions, addressees should identify the target relatively slowly. Early fixations should be equally distributed to both the target set of objects (e.g., the set of martini glasses) and the competitor set of objects (e.g., the set of jars). Looks to the target should not rise until after the onset of the object name. This condition is a baseline, since under all circumstances the latest point at which disambiguation can occur is at the object name; that is, there is always at least one object that matches the description of the intended referent, disregarding the definiteness of the referring expression. In the early disambiguation conditions, addressees should identify the target more quickly; fixations to the target set of objects should begin soon after the onset of the article and adjective, and there should be few looks to the competitor set of objects.

The crucial question was what would happen when there was a mismatch between the confederate's and the addressee's perspectives in the context of a display that would, under matching conditions, normally provide an early point of disambiguation.

Taking the speaker's perspective in these conditions requires that the addressee remember what the confederate thinks the sets of objects are like while ignoring conflicting perceptual information. If use of common ground information is delayed, then addressees should initially interpret the referring expression from their own perceptual perspective, which conflicts with what the speaker believes. Thus, there should be early looks to the competitor objects (e.g., to the single empty jar in response to *the empty*), and a delay in the identification of the intended referent. However, if addressees are able to adopt the speaker's perspective, then reference resolution should still show an advantage compared to the late disambiguation conditions.

Figure 9 also presents the proportion of fixations in each condition over the course of 2000 ms in 33 ms intervals, beginning with the onset of the referring expression.

In the matching perspective conditions, looks to the target object begin to separate from looks to other potential referents earlier for the early disambiguation condition (Figure 13a) compared to the late disambiguation condition (Figure 13b). The results demonstrated that addressees assigned reference as soon as a potential referent was uniquely identifiable given the information provided by contrast and definiteness, replicating and extending the findings of Eberhard et al. (1995), Sedivy et al. (1999), and Chambers et al. (2002).

Critically, the same pattern held with mismatching perspectives. Although addressees had to remember that the speaker thought that the objects were different than they really were, they were still able to make rapid use of information provided by

uniqueness and contrast. There was some cost associated with perspective taking, but as the comparison of Figure 13b and Figure 13c reveals, participants identified the referent faster in the mismatching condition than in the late matching condition.

In sum then, the results presented in this section confirm Keysar et al's (2000) conclusion that common ground does not completely circumscribe the referential domain for referring expressions. However, we found clear evidence that addressee's can take into account information about common ground during the earliest moments of reference resolution. Note, however, that while our results demonstrate that addressees can use limited capacity cognitive resources to consciously track a speaker's knowledge and hold onto this information in memory, this may not be the natural way that addressees keep track of the speaker's perspective. Factors such as eye-gaze, gesture, head position, and postural orientation are likely to provide cues that allow participants to track each other's perspectives, attentional states, and intentions, without requiring memory intensive cognitive models of mutual belief. Just as basic low-level cognitive mechanisms such as priming may be at least partially responsible for many phenomena that support coordination (e.g., lexical entrainment), tracking the perspective of a conversational participant might be accomplished in part via basic low-level mechanisms which social primates use to monitor each other during interaction. Whether or not these mechanisms result in fully-developed internal representations of the intentions and beliefs of a conversational partner is an important question; but, this question is orthogonal to the more basic question of how conversational partners achieve the coordination necessary for successful real-time communication.

How can we reconcile the results reported here with some of the striking demonstrations of speaker and addressee egocentricity provided by Keysar and colleagues? We have proposed that common ground can be most fruitfully viewed as a probabilistic constraint within the framework of constraint-based processing models (e.g., MacDonald, 1994; Tanenhaus & Trueswell, 1995; Trueswell & Gleitman, this volume). In constraint-based models, different information sources or constraints each contribute probabilistic evidence for alternative interpretations during processing. The constraints are weighted according to their salience and reliability, and are integrated with each other in parallel, causing the alternative interpretations to compete with each other (Spivey & Tanenhaus, 1998; McRae, Spivey-Knowlton & Tanenhaus, 1998). Factors such as speaker perspective can be incorporated into constraint-based models through expectation-based constraints, such as the likelihood that a speaker will refer to a particular entity (cf. Arnold, 2001).

Thus far, constraint-based models have been applied primarily to situations where the strength of constraints, even contextual constraints such as the prior discourse, can be estimated from relatively static experience-based factors such as frequency and plausibility. In conversational interactions in which the participants have behavioral goals, however, the state of the context must be based upon the speakers' and addressees' intentions and actions. Under these circumstances the strength and relevance of different constraints will have to be computed with respect to continuously updated contextual models because the relevancy of constraints changes moment by moment. Developing formal models of dynamically updated context will be a major

challenge for constraint-based models of comprehension, as well as for other classes of models. We should note that this challenge is similar to that faced by models of perception and action as they seek to accommodate the increasing evidence that basic perceptual processes are strongly influenced by attention and intention, which are guided by behavioral goals.

5.0. Referential domains in natural interactive conversation

Thus far we have provided empirical evidence that even the earliest moments of language comprehension are affected by factors that are central to the language-as-action view of context and are typically ignored in real-time language processing. We have not, however, established that one can actually conduct research that combines fully interactive conversation with rigorous examination of moment-by moment processing. It is important to do this for at least two reasons. First, there are likely to be differences in the behavior of participants who are participating in conversations with confederates using scripted responses compared to conversations with true interlocutors, who are generating their language on the fly. Second, there is a general methodological concern about results from experiments with scripted utterances that motivates the importance of developing complementary paradigms with non-scripted language. Despite the best efforts of experimenters to avoid creating predictable contingencies between the referential world and the form of the utterances, most controlled experiments are likely to draw attention to the feature of the language being

studied. Sentences or utterances of a particular type are likely to be over-represented in the stimulus set, and the participant's attention is likely to be drawn to small differences in structure or usage that are being investigated in the experiment. Thus it would be desirable to replicate results from experiments with scripted utterances using more natural spontaneous utterances.

In a recent study (Brown-Schmidt, Campana & Tanenhaus, in press), we used a modified version of a referential communication task to investigate comprehension of definite referring expressions, such as *the red block*. Pairs of participants, separated by a curtain, worked together to arrange blocks in matching configurations and confirm those configurations. The characteristics of the blocks afforded comparison with findings from scripted experiments investigating language-driven eye movements, specifically those demonstrating point of disambiguation effects during reference resolution. We investigated: (1) whether these effects could be observed in a more complex domain during unrestricted conversation, and (2) under what conditions the effects would be eliminated, indicating that factors outside of the speech itself might be operating to circumscribe the referential domain.

Figure 10 about here

Figure 10 presents a schematic of the experimental setup. In order to encourage participants to divide up the workspace (e.g. the board) into smaller referential domains, we divided participants' boards into five physically distinct sub-areas. Initially, sub-areas contained stickers representing blocks. The task was to replace each sticker with a matching block. While partners' boards were identical with respect to

sub-areas, partners' stickers differed: where one partner had a sticker, the other had an empty spot. Pairs were instructed to tell each other where to put blocks so that in the end their boards would match. No other restrictions were placed on the interaction. The entire experiment lasted approximately 2.5 hours. For each pair we recorded the eye movements of one partner and the speech of both partners.

The stickers (and corresponding blocks) were of two types. Most of the blocks were of assorted shapes (square or rectangle) and colors (red, blue, green, yellow, white or black). The initial configuration of the stickers was such that the color, size, and orientation of the blocks would encourage the use of complex noun phrases and grounding constructions. Some of the stickers (and corresponding blocks) contained pictures which could be easily described by naming the picture (e.g. 'the candle'). We selected pairs of pictures that referred to objects with initially acoustically overlapping names, cohort competitors such a clown and cloud. Half of the cohort competitor stickers were arranged such that both cohort competitor blocks would be placed in the same sub-area of the board. The other half of the cohort competitor stickers were arranged such that the cohort competitor blocks would be placed in different sub-areas of the board. All of the cohort competitor pairs were separated by approximately 3.5 inches.

The conversations for each of the four pairs were transcribed, and eye movements associated with definite references to blocks were analyzed. The non-eye tracked partners generated a total of 436 definite references to colored blocks. An analysis of these definite references demonstrated that just over half (55%) contained a

linguistic point of disambiguation, while the remaining 45% were technically ambiguous with respect to the sub-area that the referent was located in (e.g. 'the red one' uttered in a context of multiple red blocks). Two researchers coded the noun phrases for their point of disambiguation (POD), defined as the onset of the word in the noun phrase uniquely identified a referent, given the visual context at the time. Average POD was 864ms following the onset of the noun phrase. Eye movements elicited by noun phrases with a unique linguistic point of disambiguation were analyzed separately from those that were never fully disambiguated linguistically. The eye-tracking analysis was restricted to cases where at least one competitor block was present. This left 74 linguistically disambiguated trials and 192 ambiguous trials.

Figure 11 about here

Eye movements elicited by disambiguated noun phrases are pictured in Figure 11a. Before the POD, subjects showed a preference to look at the target block. Within 200ms of the onset of the word in the utterance that uniquely specified the referent, looks to targets rose substantially. This point of disambiguation effect for looks to the target is similar to that seen by Eberhard, et al. (1995), demonstrating that we were successful in using a more natural task to investigate on-line language processing. The persistent target bias and lack of a significant increase in looks to competitors are likely due to additional pragmatic constraints that we will discuss shortly.

Most remarkably, while we found clear point of disambiguation effects for disambiguated noun phrases, for ambiguous utterances (see Figure 11b) fixations were primarily restricted to the referent. Thus the speaker's underspecified referential expressions did not confuse listeners, indicating that referential domains of the speaker and the listener were closely coordinated. These results suggest that (1) speakers systematically use less specific utterances when the referential domain has been otherwise constrained; (2) the attentional states of speakers and addresses become closely coordinated? ; and (3) utterances are interpreted with respect to referential domains circumscribed by contextual constraints.

In order to identify what factors led speakers to choose underspecified referring expressions, and enabled addresses to understand them, we performed a detailed analysis of all of the definite references, focusing on factors that seemed likely to be influencing the generation and comprehension of referential expressions. Our general hypothesis was that speakers would choose to make a referential expression more specific when the intended referent and at least one competitor block were each salient in the relevant referential domain.

We focused on recency, proximity and compatibility with task constraints. These factors are similar to those identified by Beun and Cremers (1998) using a 'construction' task in which participants, separated by a screen, worked together in a mutually co-present visual space to build a structure out of blocks.

Recency. We assumed that recency would influence the saliency of a referent, with the most recently mentioned entities being more salient than other (non-focused)

entities. Thus, how recently the target block was last mentioned should predict the degree of specification, with references to the most recently mentioned block of a type, resulting in ambiguous referring expressions. For example, if *the green block* were uttered in the context of a set of 10 blocks, 2 of which were green, recency would predict that the referent should be the green block that was most recently mentioned.

Proximity. We examined the proximity of each block to the last mentioned block, because partners seemed to adopt a strategy of focusing their conversation on small regions within each sub-area. In the following segment of the discourse, we see an example where the referent of an otherwise ambiguous noun phrase is constrained by proximity.

> 2. ok, so it's four down, you're gonna go over four, and then you're gonna put the piece right there
>
> 1. ok…how many spaces do you have between this green piece and the one to the left of it, vertically up?

The underlined referring expression is ambiguous given the visual context; there are approximately three green blocks up and to the left of the previously focused block (the one referred to in the NP as *this green piece*). In this case the listener does not have difficulty dealing with the ambiguity because he considers only the block closest to the last mentioned block.

Task compatibility. Task compatibility refers to constraints on block placement due to the size and shape of the board, as well as the idiosyncratic systems that partners used to complete the task. In the following exchange, compatibility circumscribes the referential domain as the participants work to agree where the clown block should be placed:

1. ok, you're gonna line it up… it's gonna go <pause> one row ABOVE the green one, directly next to it.

2. can't fit it

1. cardboard?

2. can't yup, cardboard

1. well, take it two back

2. the only way I can do it is if I move, alright, should the green piece with the clown be directly lined up with thuuh square?

Again, the underlined referring expression is ambiguous given the visual context. While the general task is to make their boards match, the current sub-task is to place the clown piece (which they call *the green piece with the clown*). In order to complete this sub-task, Speaker 2 asks whether the clown should be lined up with the target, *thuuh square*. The listener does not have difficulty dealing with this ambiguous reference because, although there are a number of blocks one could line up with *the green piece with the clown*, only one is task-relevant. Given the location of all the blocks

in the relevant sub-area, the target block is the easiest block to line up with the clown. The competitor blocks are inaccessible because of the position of the other blocks or the design of the board.

For all ambiguous and disambiguated trials, each colored block in the relevant sub-area was coded for recency (number of turns since last mention), proximity (ranked proximity to last mentioned item) and task constraints (whether or not the task predicted a reference to that block). Targets consistently showed an advantage for all three constraints, establishing their validity

Planned comparisons revealed that target blocks were more recently mentioned and more proximal than competitor blocks, additionally target blocks best fit the task constraints. However, recency, proximity, and task compatibility of the target blocks did not predict speaker ambiguity. Instead, speaker ambiguity was determined by the proximity and task constraints associated with the competitor blocks. When a competitor block was proximate and fit the task constraints, speakers were more likely to linguistically disambiguate their referential expression. A logistic regression model supported these observations: noun phrase ambiguity was significantly predicted by a model which included Task and Proximity effects, with no independent contribution of Recency.

These results demonstrate that the relevant referential domain for the speakers, and addressees was restricted to a small task-relevant area of the board. Striking support for this conclusion comes from an analysis of trials in which there was a cohort competitor for the referent in the addressee's referential domain. When referential

domains are unrestricted, listeners typically look at cohort competitors more often than distracters with unrelated names (Allopenna et al., 1998; Dahan, Magnuson & Tanenhaus, 2001). Moreover, fixations generated while the word is still ambiguous are equally likely to be to the cohort competitor as to the target. However, Brown-Schmidt et al. found that looks to cohort competitors were no more likely than looks to competitors with unrelated names. This is not simply a null effect. Because of the length of the experiment, participants occasionally needed to take a bathroom break. Following the break, the eye tracker had to be recalibrated. The experimenter did so by instructing the participant to look at some of the blocks. Under these conditions, the referential domain consists of the entire display because there is no constraining conversational or task-based goal. When the intended referent had a cohort competitor, the participant also frequently looked at the competitor, showing the classic cohort effect.

In sum, these results demonstrate that it is possible to study real-time language processing in a complex domain during unrestricted conversation. When a linguistic expression is temporarily ambiguous between two or more potential referents, reference resolution is closely time-locked to the word in the utterance that disambiguates the referent, replicating effects found in controlled experiments with less complex displays and pre-scripted utterances. Most importantly, our results provide a striking demonstration that participants in a task-based or "practical dialogues" (Allen et al., 2001), closely coordinate referential domains as the conversation develops. Speakers chose referential expressions with respect to a circumscribed task-relevant

referential domain. Speakers were more likely to choose a specific referential expression when both the referent and a competitor were salient in the immediate task-relevant environment (saliency of competitors was predicted by task and proximity constraints). When only the referent was salient, speakers were less likely to add additional modification, generating utterances whose referents would be ambiguous if one took into account only the local visual context and the utterance. This by itself makes sense and is unremarkable, though the degree of ambiguity is perhaps surprising. What is remarkable, however, is that listeners were rarely even temporarily confused by underspecified referential expressions. This demonstrates that the participants in the conversation had developed closely matching referential domains, suggesting that referential domains become closely aligned as proposed by Pickering and Garrod (in press). Moreover, reference resolution appeared to be affected by collaborative constraints that developed during the conversation. Our participants spontaneously created collaborative terms for troublesome words (such as *horizontal* and *vertical*), and tuned their utterances, and comprehension systems for such details as the recency of mention of each particular kind of block, proximity of blocks to one another, and task constraints idiosyncratic to our block-game. These observations suggest that the attentional states of the interlocutors become closely tuned during the course of their interaction.

6.0 Summary and implications

We began this chapter by noting that most research on language comprehension has been divided into two distinct traditions, the language-as-product and the language-as-action traditions, each with its own conception of what comprise the fundamental problems in language processing, and each with its own preferred methodologies. The product tradition emphasizes the cognitive processes by which listeners recover linguistic representations in real-time using response measures that are closely time-locked to the linguistic stimuli. The action tradition emphasizes the processes by which interlocutors cooperate to achieve joint goals, using natural tasks, typically with real-world referents and well-defined behavioral goals. We also argued that these traditions have conflicting views of what constitutes the context for comprehension. For the product tradition, context is one of many factors that might modulate core processes that are fundamentally context independent, whereas context is intrinsic to language processing in the action tradition. We then asked whether the notion of context espoused by the action tradition affects the earliest moments of real-time language processing, focusing on the referential domain for definite referring expressions. We showed that actions, intentions, real-world knowledge, and mutual knowledge circumscribe referential domains, and that these context-specific domains affect core processes, such as syntactic ambiguity resolution.

In order to appreciate how this notion of context-specific referential domain alters the standard product-based view of real-time sentence processing, consider the utterance in:

5. After John put the pencil below the big apple, he put the apple on top of the towel.

A traditional account goes something like this. When the listener encounters the scalar adjective *big,* interpretation is delayed because a scalar dimension can only be interpreted with respect to the noun it modifies: compare, for example, a big building and a big pencil. As *apple* is heard, lexical access activates the apple concept, a prototypical apple. The apple concept is then modified resulting in a representation of a BIG APPLE. When *apple* is encountered in the second clause, lexical access again results in activation of a prototypical APPLE concept. Because *apple* was introduced by a definite article, this representation would need to be compared with the memory representation of the BIG APPLE to decide whether the two corefer. (See, for example, Tanenhaus, Carlson & Seidenberg, 1985, for an outline of such an approach.)

This account of moment-by-moment processing seems reasonable when we focus on the processing of just the linguistic forms. However, let's reconsider how the real time interpretation of (5) proceeds in the context illustrated in Figure 12, taking into account results we have reviewed. At *big*, the listener's attention will be drawn to the larger of the two apples, because a scalar adjective signals a contrast among two or more entities of the same semantic type. Thus *apple* will be immediately interpreted as the misshapen green apple, despite the fact that there is a more prototypical red apple in the scene. When *the apple* is encountered in the second clause, it will be immediately interpreted as the large green apple, despite the fact that there is also a co-present prototypical red apple in the scene. Moreover, *put* would be interpreted as a specific

action, involving either John's hand or, if John was holding an instrument, an action using the instrument, and the affordances it introduced.

Figure 12 about here

A common product-based view is to construe the goal of real-time sentence processing research as determining how listeners compute context-independent representations. The output of this "input" system will be a linguistic representation that can be computed quickly because the system is encapsulated. One of the appealing aspects of context-independent representations is that they could be computed by fast, limited-domain, processing systems. Moreover, such representations would be general enough to serve as the skeleton for more computationally complex context-specific representations. Now, however, we see that people can, and do, map linguistic input onto action-based representations from the earliest moments of processing. Moreover, we find increasing evidence throughout the brain and cognitive sciences that (a) behavioral context, including attention and intention affect even basic perceptual processes (e.g., Gandhi, Heeger, & Boyton, 1998; Colby & Goldberg, 1999) and (b) brain systems involved in perception and action are implicated in the earliest moments of language processing (e.g., Pulvermüller, Härle, & Hummel, 2001). Given these developments, focusing solely on context-independent, form-based processing is unlikely to prove fruitful (cf., Barsalou, 1999; Glenberg & Robertson, 2000; Spivey, Richardson & Fitneva, this volume). Thus an important goal for future research will be integrating action-based notions of linguistic context with perceptual and action-based

accounts of perception, language and cognition in order to develop more realistic models of real-time language processing.

This perspective appears to be increasingly shared by researchers in related fields of study. For example, as work in theoretical linguistics continues to marshal evidence that even basic aspects of semantic interpretation involve pragmatic inference (e.g., Levinson, 2000), "embodied" approaches to discourse understanding have concurrently demonstrated that semantic representations incorporate perceptual factors and knowledge of how referents are acted upon in specific situations (e.g., Barsalou, 1999; Glenberg & Robertson, 2000; also Altmann & Kamide, this volume; Spivey, et al., this volume). In addition, there is now substantial evidence that social pragmatic cues such as joint attention and intentionality are critical in early language development (e.g., Bloom, 1997; Sabbagh & Baldwin, 2001), as well as evidence showing that nonlinguistic gestures contribute to the understanding of speech (e.g., Goldin-Meadow, 1999; McNeill, 2000). By moving towards a methodological and theoretical union of the action and product traditions, work in language processing can more fruitfully identify points of contact with these areas of research.

Acknowledgments

*This work was partially supported by NIH grants HD-27206 and NIDCD DC-005071 to MKT, NSERC 69-6157 to CGC, and NRSA MH1263901A1 to JEH. We thank John Henderson, Fernanda Ferreira, Zenzi Griffin, and Sarah Brown-Schmidt for helpful comments.

Footnotes

1. This conclusion depends on the assumption that behavioral data about syntactic preferences in parsing bear on questions about the architecture of the language processing system. Thus, just as the absence of context effects is claimed to be evidence for modularity, the presence of context effects provides evidence against modularity (Fodor, 1983, Frazier, 1999). Note, however, that there is an alternative view in which behavioral data reflect the output of rapid general-purpose cognitive decisions that act upon the output of a modular parallel parser. With apologies to William James, we will refer to this as "turtles" modularity because it assumes that an unobservable process underlies a second observable process.

References

Allen, J. F., Byron, D. K., Dzikovska, M., Ferguson, G., Galescu, L., & Stent, A. (2001). Towards conversational human-computer interaction. *AI Magazine*. **22**, 27-35.

Allopenna, P. D, Magnuson, J.S. & Tanenhaus, M.K. (1998). Tracking the time course of spoken word recognition: evidence for continuous mapping models. *Journal of Memory and Language*, **38**, 419-439.

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

Altmann, G. T. M. & Kamide, Y. this volume.

Altmann, G.T.M., & Steedman, M.J. (1988). Interaction with context during human sentence processing. *Cognition,* **30**, 191-238.

Arnold, J. E. (2001). The effect of thematic roles on pronoun use and frequency of reference continuation. *Discourse Processes,* **31**, 137-162.

Arnold, J.E., Trueswell, J.C., & Lawentmann, S.M. (1999, November). Using common ground to resolve referential ambiguity. Paper presented at the 40th Annual Meeting of the Psychonomic Society. Los Angeles, CA.

Austin, J.L. (1962). *How to do things with words*. Oxford: Oxford University Press.

Barsalou, L. (1999). Language comprehension: Archival memory or preparation for situated action? *Discourse Processes*, **28**, 61-80.

Barsalou, L. (1999). Language comprehension: Archival memory or preparation for situated action? *Discourse Processes*, **28**, 61-80

Beun, R.-J. and Cremers, A.H.M. (1998). Object reference in a shared domain of conversation. *Pragmatics & Cognition*, **6**, 121-152.

Bloom, P. (1997). Intentionality and word learning. *Trends in Cognitive Sciences*, **1**, 9-12.

Bloom, P. (1997). Intentionality and word learning. *Trends in Cognitive Sciences*, **1**, 9-12.

Brown-Schmidt, S., Campana, E. & Tanenhaus, M.K. (in press). Real-time reference resolution in a referential communication task. In J.C. Trueswell & M.K. Tanenhaus, (Eds). *Processing world-situated language*: *Bridging the language-as-action and language-as-product traditions*. Cambridge, MA: MIT Press.

Chambers, C.G. (2001). The dynamic construction of referential domains. Unpublished Doctoral Dissertation, University of Rochester.

Chambers, C.G., Tanenhaus, M.K., & Magnuson, J.S. (2003). Action-based inference and syntactic ambiguity resolution: A test of the modularity hypothesis. Manuscript under review.

Chambers, C.G., Tanenhaus, M.K., Eberhard, K.M., Carlson, G.N. & Filip, H. (2002). Circumscribing referential domains during real-time language comprehension. *Journal of Memory and Language, 47*, 30-49.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Clark, H. H. (1992) *Arenas of Language Use*. Chicago, IL: University of Chicago.

Clark, H.H. (1996). *Using Language*. Cambridge, UK: Cambridge University Press.

Clark, H.H. (1997). Dogmas of understanding. *Discourse Processes*, **23**, 567-598.

Clark, H.H. & Brennan, S.E. (1989). Grounding in communication. In L. Resnick, J. Levine, & S. Teasley (Eds.), *Perspectives on socially shared cognition*. American Psychological Association: Washington, D.C.

Clark, H.H. and Carlson, T. (1981). Context for comprehension. In J. Long and 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.H. Joshi, B. Webber, & I.A. Sag (Eds.), *Elements of discourse understanding* (pp.10-63). Cambridge, UK: Cambridge University Press.

Clark, H.H. & Wilkes-Gibbs (1986). Referring as a collaborative process. *Cognition*, **22**, 1-39.

Colby, C.L. & Goldberg, M.E. (1999). Space and attention in parietal cortex. *Annual Review of Neuroscience*, **22**, 97-136.

Coltheart, M. (1999). Modularity and cognition. *Trends in Cognitive Sciences,* **3**, 115-120.

Cooper, R. M. (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.

Crain, S., & Steedman, M. (1985). On not being led up the garden path: the use of context by the psychological parser. In D. Dowty, L. Karttunen, & A. Zwicky (Eds.), *Natural Language Parsing: Psychological, Computational, and Theoretical Perspectives* (pp. 320-358). Cambridge, UK: Cambridge University Press.

Dahan, D., Magnuson, J.S. & Tanenhaus, M.K. (2001). Time course of frequency effects in spoken word recognition: evidence from eye movements. *Cognitive Psychology,* **42**, 317-367.

Eberhard, K.M., Spivey-Knowlton, M.J., Sedivy, J.C. & Tanenhaus, M.K. (l995). Eye-movements as a window into spoken language comprehension in natural contexts. *Journal of Psycholinguistic Research*, **24**, 409-436.

Epstein, R. (1998). Reference and definite referring expressions. *Pragmatics & Cognition*, **6**, 189-207

Fodor (1983). *Modularity of mind*. Cambridge, MA: Bradford Books.

Frazier, L. (1999). Modularity and language. In R.A. Wilson and F.C. Keil (Eds.) *MIT encyclopedia of cognitive science*, (pp. 557-558). Cambridge, Mass: MIT Press.

Gandhi, S.P., Heeger, M.J. & Boyton, G.M. (1998). Spatial attention affects brain activity in human primary visual cortex. *Proceedings of the National Academy of Science USA* **96**, 3314-3319.

Glenberg, A.M., & Robertson, D.A. (2000). Symbol grounding and meaning: A comparison of high-dimensional and embodied theories of meaning. *Journal of Memory and Language,* **43**, 379-401.

Goldin-Meadow, S. (1999). The role of gesture in communication and thinking. *Trends in Cognitive Sciences*, **3**, 419-429.

Grice, H. P. (1957). Meaning. *Philosophical Review* **66**, 377-388.

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

Horton, W.S. & Keysar, B. (1995). When do speakers take into account common ground? *Cognition*, **59**, 91-117.

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**, 32-37.

Keysar, B., Barr, D.J., & Horton, W.S. (1998). The egocentric basis of language use: Insights from a processing approach. *Current Directions in Psychological Science*, **7,** 46-50.

Krauss, R. M. and Weinheimer, S. (1966) Concurrent feedback, confirmation, and the encoding of referents in verbal communication. *Journal of Personality and Social Psychology*, **4**, 343-346.

Levinson, S.C. (2000). *Presumptive Meanings*. Cambridge, MA: MIT Press.

MacDonald, M.C. (1994). Probabilistic constraints and syntactic ambiguity resolution. *Language and Cognitive Processes*, **9**, 157-201.

McMurray, B., Tanenhaus, M.K.& Aslin, R.N. (2002). Gradient effects of within-category phonetic variation on lexical access. *Cognition*, **86**, B33-42.

McNeill, D. (Ed). (2000). *Language and Gesture*. Cambridge, UK: Cambridge University Press.

McRae, K., Spivey-Knowlton, M. J., & Tanenhaus, M. K. (1998). Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension. *Journal of Memory and Language*, **38**, 283-312.

Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology*, **90**, 227-234.

Miller, G. A. (1962). Some psychological studies of grammar. *American Psychologist*, **17**, 748-762.

Miller, G. A. & Chomsky, N. (1963). Finitary models of language users. In R. D. Luce, R. R. Bush, & E. Galanter, (Eds.), *Handbook of mathematical psychology*. New York: Wiley.

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

Pickering, M. & Garrod, S.A. (in press). Towards a mechanistic psycholinguistics of dialogue. *Brain and Behavioral Sciences*.

Pulvermüller, F., Härle, M., & Hummel, F. (2001). Walking or talking? Behavioral and neurophysiological correlates of action verb processing. *Brain and Language*, **78**, 143-168.

Rayner, K.E. & Liversedge, S.P. this volume

Sabbagh, M.A. & Baldwin, D.A. (2001). Learning words from knowledgeable versus ignorant speakers: Links between preschoolers' theory of mind and semantic development. *Child Development*, **72**, 1054-1070.

Schegloff, E. A & Sacks, H. (1973). Opening up closings. *Semiotica*, **8**, 289-327.

Searle, J.R. (1969). *Speech acts. An essay in the philosophy of language*. Cambridge: Cambridge University Press.

Sedivy, J.C., Tanenhaus, M.K., Chambers, C.G. & Carlson, G.N. (1999). Achieving incremental processing through contextual representation: Evidence from the processing of adjectives. *Cognition*, **71**, 109-147.

Spivey,M.J., Richardson, D.C. & Fitenva, S. A. this volume.

Spivey, M. J., & Tanenhaus, M. K. (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.

Spivey, M.J., Tanenhaus, M.K., Eberhard, K.M. & Sedivy, J.C. (2002). Eye movements and spoken language comprehension: Effects of visual context on syntactic ambiguity resolution. *Cognitive Psychology*, **45**, 447-481.

Swinney, D.A., Onifer, W., Prather, P. & Hirshkowitz, M. (1978). Semantic facilitation across sensory modalities in the processing of individual words and sentences. *Memory and Cognition*, **7**, 165-195.

Tanenhaus, M.K., Carlson, G. & Seidenberg, M.S. (1985). Do listeners compute linguistic representations? In D. Dowty, L. Kartunnen & A. Zwicky (Eds.), *Natural Language Parsing: Psychological, Computational, and Theoretical Perspectives*, (pp.359-408). Cambridge: Cambridge University Press.

Tanenhaus, M. K., Magnuson, J. S., Dahan, D., & Chambers, C. G. (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.

Tanenhaus, M.K., Spivey-Knowlton, M.J., Eberhard, K.M., & Sedivy, J.C. (1995). Integration of visual and linguistic information during spoken language comprehension. *Science*, **268**, 1632-1634.

Tanenhaus, M.K., & Trueswell, J.C. (1995). Sentence comprehension. In J. Miller & P. Eimas (Eds.), *Speech, Language, and Communication* (pp. 217-262), San Diego, CA: Academic Press.

Trueswell, J.C. & Gleitman, L.A. this volume

Trueswell, J.C., Sekerina, I., Hill, N. & Logrip, M. (1999). The kindergarten-path effect: Studying on-line sentence processing in young children. *Cognition*, **73**, 89-134.

Figure Captions

Figure 1. Schematic of prototypical product experiment: cross-modal lexical priming with lexical decision

Figure 2. Schematic of prototypical action experiment: referential communication task with Tangrams.

Figure 3. The top panel shows sample stimuli. The small cube will fit into both cans, but the large cube will fit into only the big can. The bottom panel shows the mean latency to launch an eye movement to the goal with definite and indefinite instructions and one and more than one compatible goal referents.

Figure 4. The top panel shows sample stimuli for one-referent (pencil) and two-referent (apple on napkin) conditions. The bottom panel shows the proportion of looks to the competitor goal (the towel) for instructions with locally ambiguous and unambiguous prepositional phrases in one-referent and two-referent contexts.

Figure 5. The top panel shows sample stimuli for trials with action-compatible competitor (liquid egg in glass) and action-incompatible competitor (solid egg in glass). The bottom panel shows the mean proportion of time spent looking at the competitor

goal (the empty bowl) for instructions with locally ambiguous and unambiguous prepositional phrases with action-compatible and action-incompatible competitors

Figure 6. The top panel shows sample stimuli. Both whistles can be moved by hand, but only the whistle with the string attached can be picked up with a hook. The bottom panel shows the proportion of looks to the competitor goal when the presence of absence of an instrument makes the competitor action-compatible or action-incompatible.

Figure 7. Example displays and critical instruction for a single item rotated through the conditions in which the competitor was in common ground (left column) or privileged ground (right column), and when it was the same color (top row) or a different color (bottom row). All of the shapes on the board were known to both the confederate speaker and the participant addressee except for the secret shape, indicated here with a gray background, which was only known to the addressee. The target shape location in this display was the topmost red (R) triangle.

Figure 8. Proportion of fixations for each object type over time in the common ground with a different color competitor condition (upper graph) and the privileged ground with a different color competitor condition (lower graph).

Figure 9. Left panels show sample displays for the late disambiguation/matching perspective, early disambiguation/matching perspective, and mismatching perspective conditions. Right panels show proportions of fixations on each object type over time. Points of disambiguation in the instructions are indicated in bold.

Figure 10. Schematic of the setup for the referential communication task. Solid regions represent blocks; striped regions represent stickers (which will eventually be replaced with blocks). The scene pictured is midway through the task, so some portions of the partners' boards match, while other regions are not completed yet.

Figure 11. The top graph shows the proportion of fixations to targets, competitors, and other blocks by time (ms) for linguistically disambiguated definite noun phrases. The graph is centered by item with 0 ms = POD onset. The bottom graph shows the proportion of fixations for the linguistically ambiguous definite noun phrases.

Figure 12. Hypothetical context for utterance, *After John put the pencil below the big apple, he put the apple on top of the towel*, to illustrate the implausibility of some standard assumptions of context-independent comprehension. Note that the small (red) apple is intended to be a more prototypical apple than the large (green) apple.

Figure 1. Schematic of prototypical product experiment: cross-modal lexical priming with lexical decision

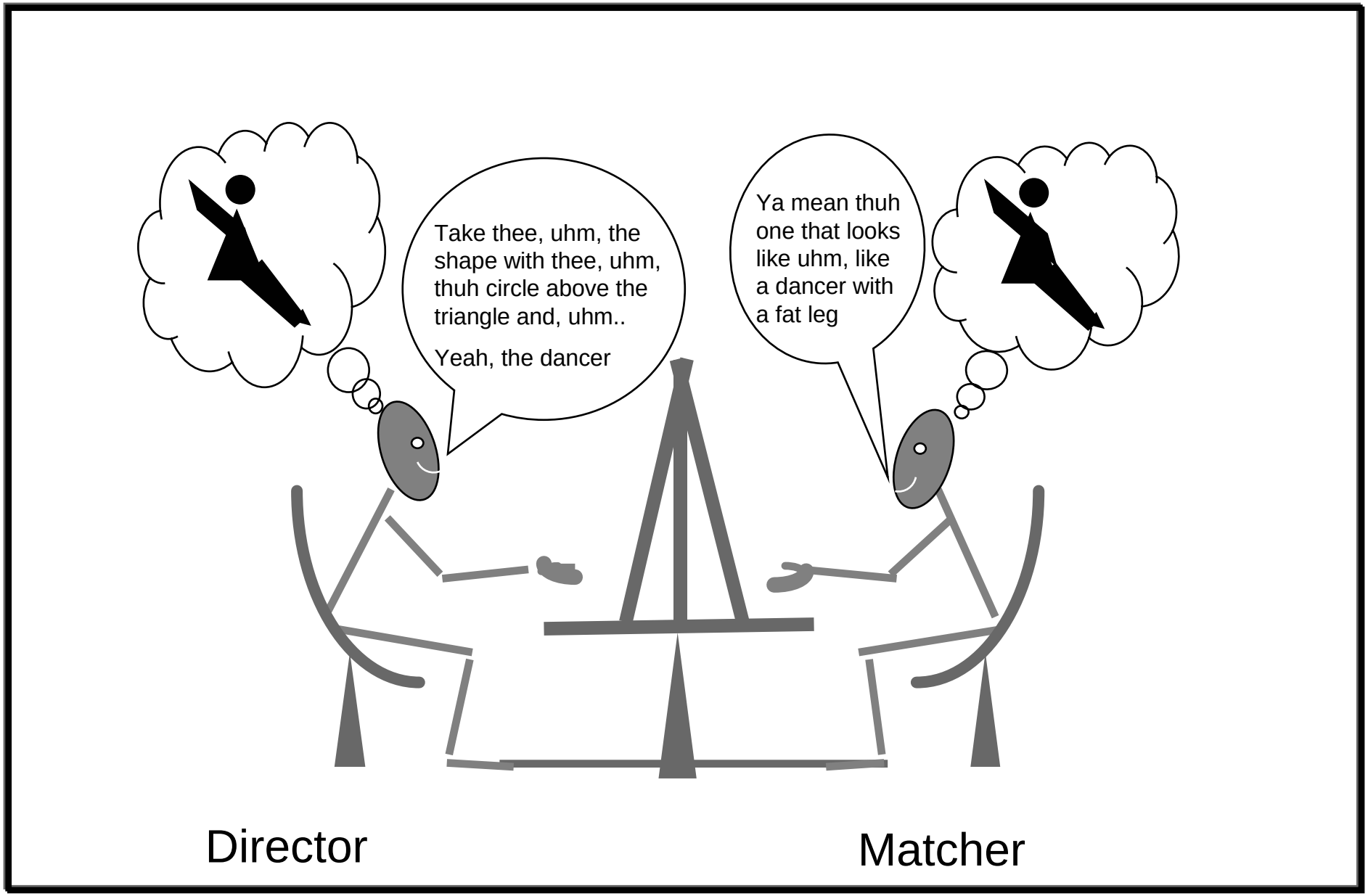
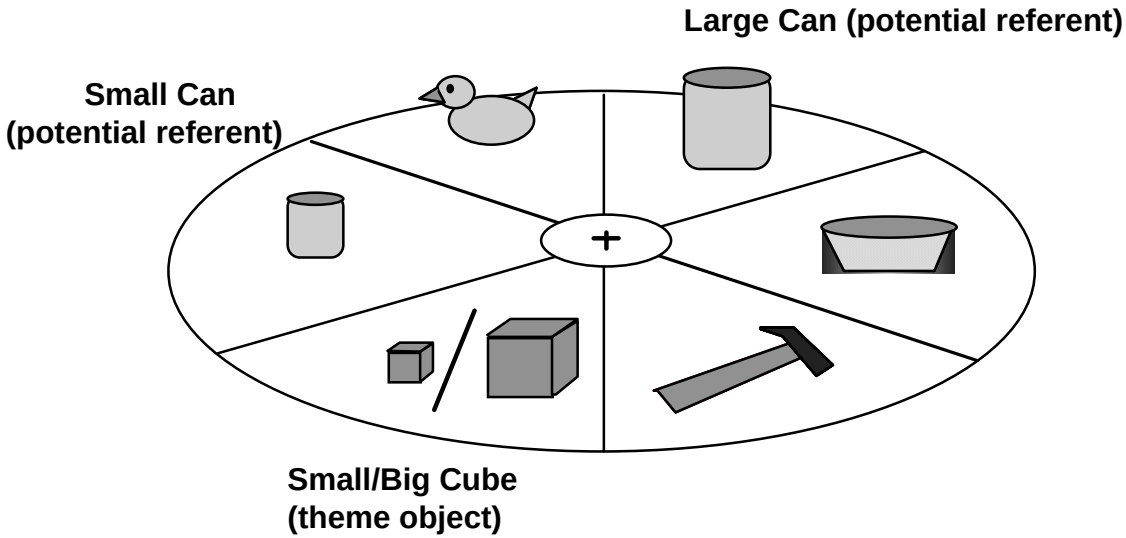


Figure 2. Schematic of prototypical Action task: referential communication task with Tangrams,



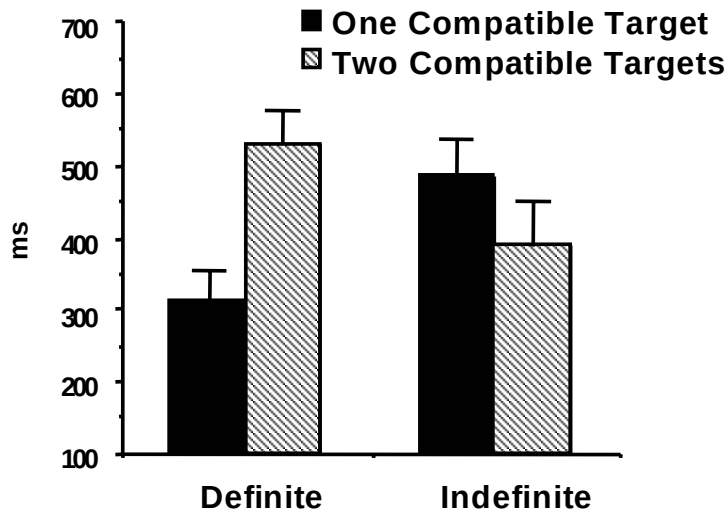


Figure 3. The top panel shows sample stimuli. The small cube will fit into both cans, but the large cube will fit into only the big can. The bottom panel shows the mean latency to launch an eye movement to the goal with definite and indefinite instructions and one and more than one compatible goal-referents.

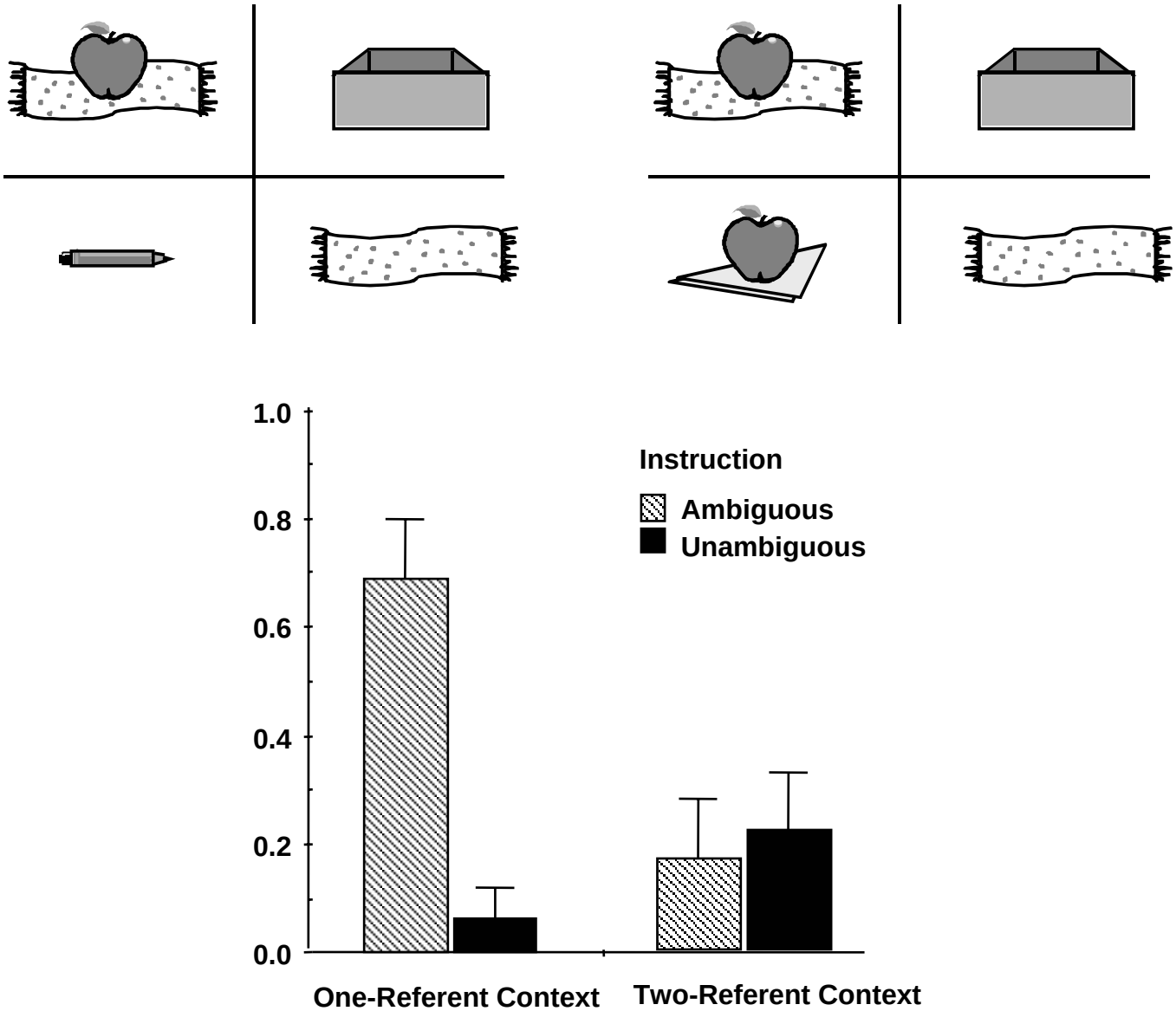


Figure 4. The top panel shows sample stimuli for one-referent (pencil) and two-referent (apple on napkin) conditions. The bottom panel shows the proportion of looks to the competitor goal (the towel) for instructions with locally ambiguous and unambiguous prepositional phrases in one-referent and two-referent contexts.

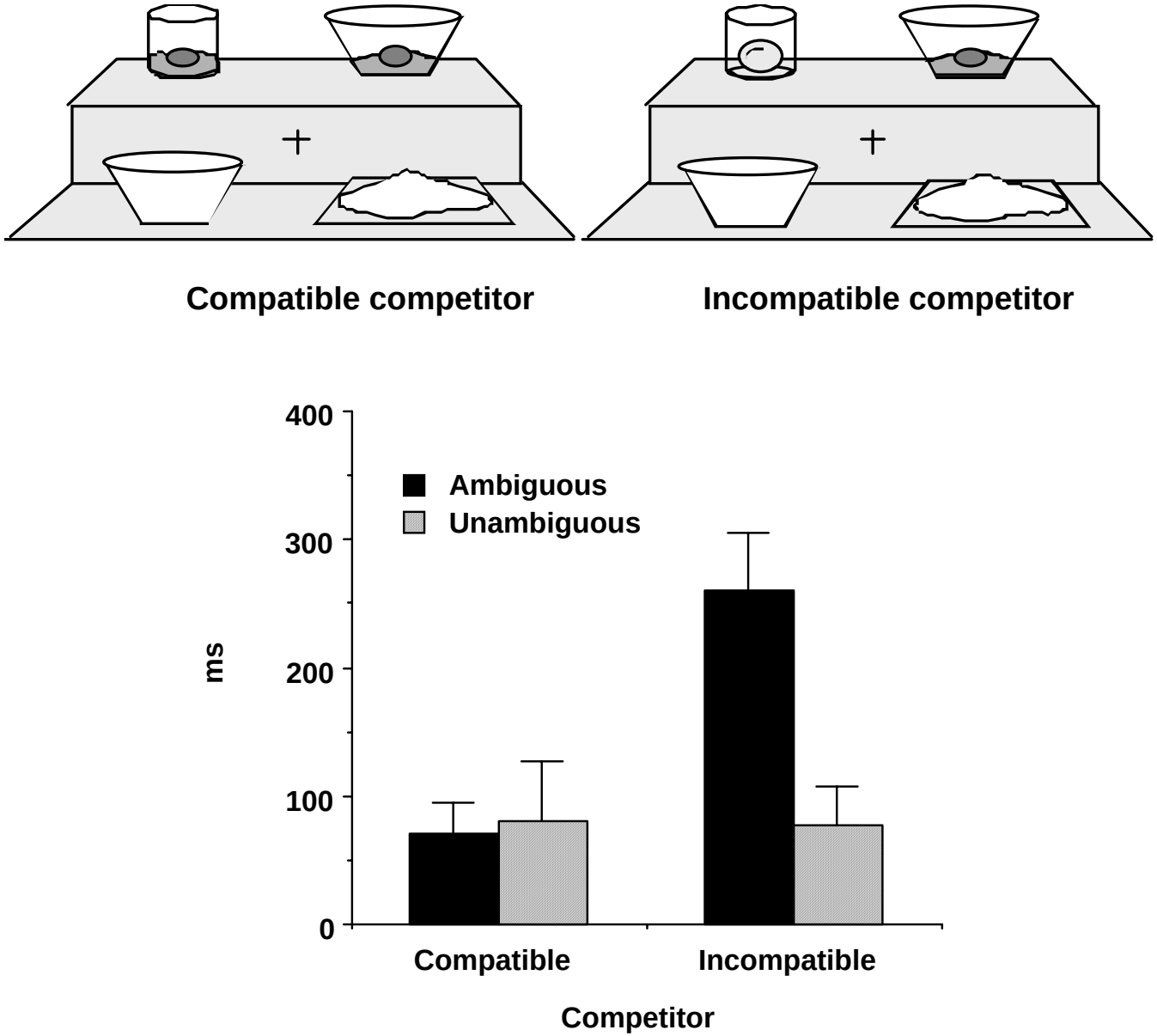


Figure 5. The top panel shows sample stimuli for trials with action-compatible competitor (two liquid eggs) and action-incompatible competitor (one solid egg). The bottom panel shows the mean proportion of time spent looking at the competitor goal (the empty bowl) for instructions with locally ambiguous and unambiguous prepositional phrases with action-compatible and action-incompatible competitors.

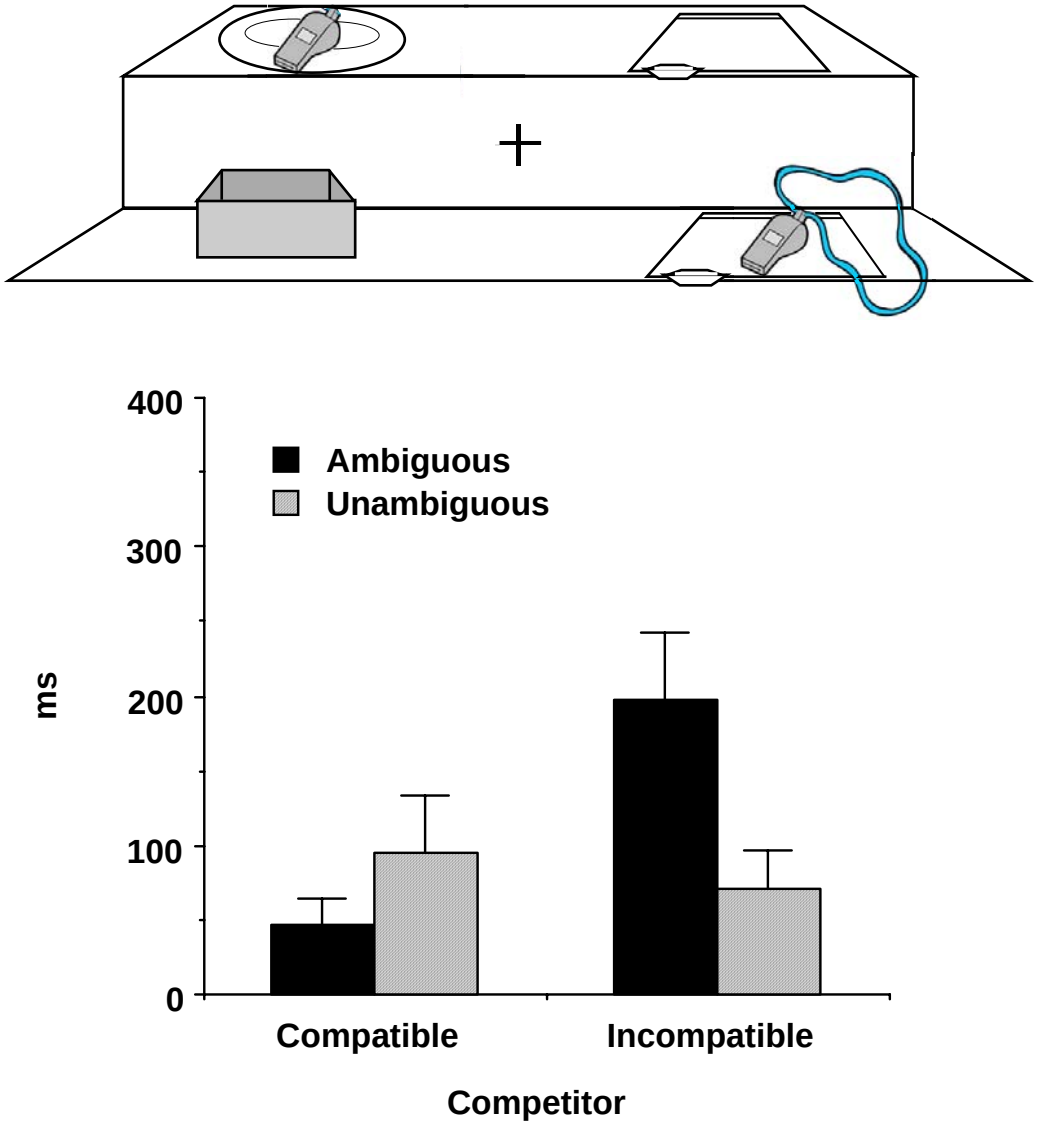
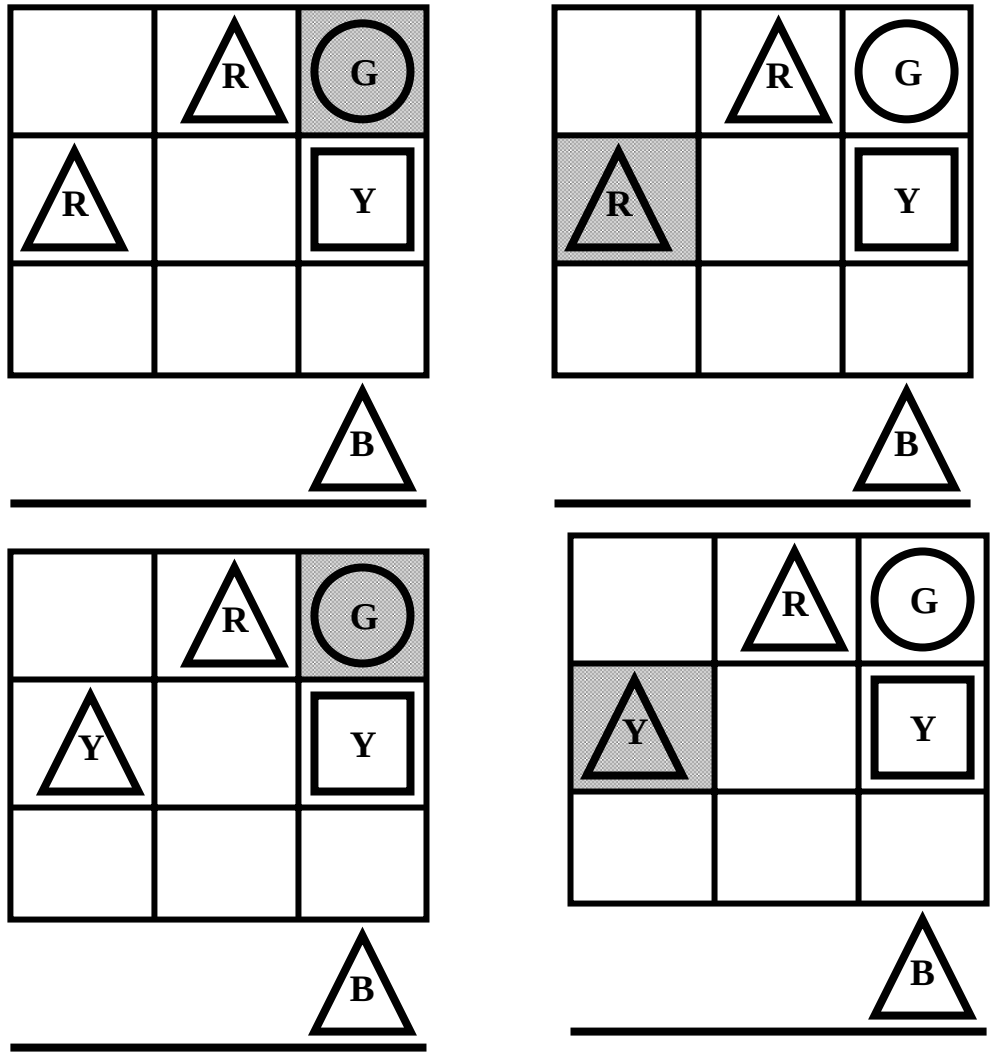


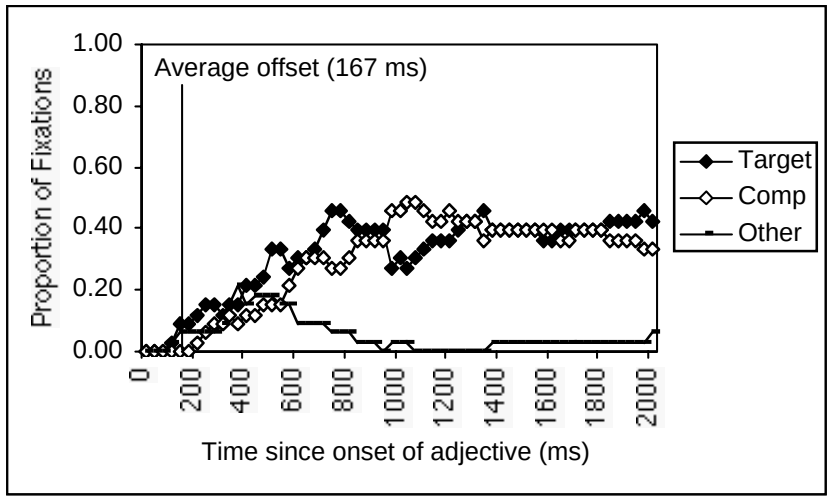
Figure 6. The top panel shows sample stimuli. Both whistles can be moved by hand, but only the whistle with the string attached can be picked up with a hook. The bottom panel shows the proportion of time spent looking at the competitor goal when the presence or absence of an instrument makes the competitor action-compatible or action-incompatible.



Now put the blue triangle on <u>the red one</u>.

Figure 7. Example displays and critical instruction for a single item rotated through the conditions in which the competitor was in common ground (left column) or privileged ground (right column), and when it was the same color (top row) or a different color (bottom row). All of the shapes on the board were known to both the confederate speaker and the participant addressee except for the secret shape, indicated here with a gray background, which was only known to the addressee. The target shape location in this display was the topmost red (R) triangle.

a. Common Ground/Same Color Competitor



b. Privileged Ground/Same Color Competitor

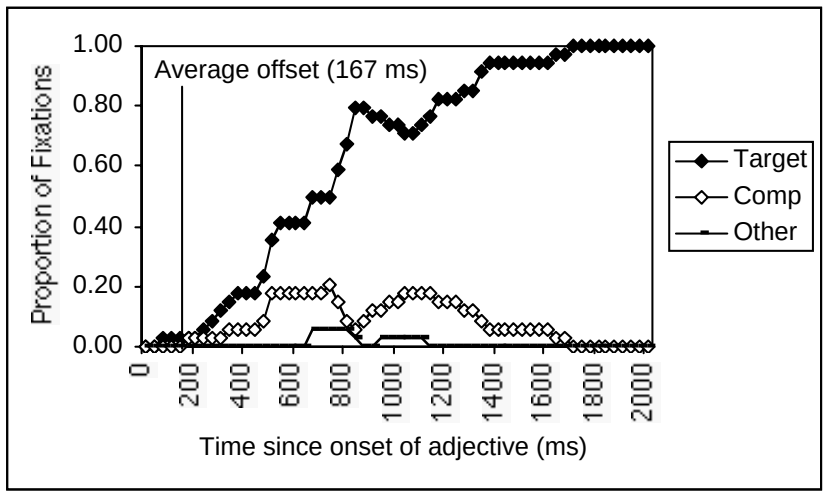


Figure 8. Proportion of fixations for each object type over time in the common ground/different color competitor condition (upper graph) and the privileged ground/different color competitor condition (lower graph).

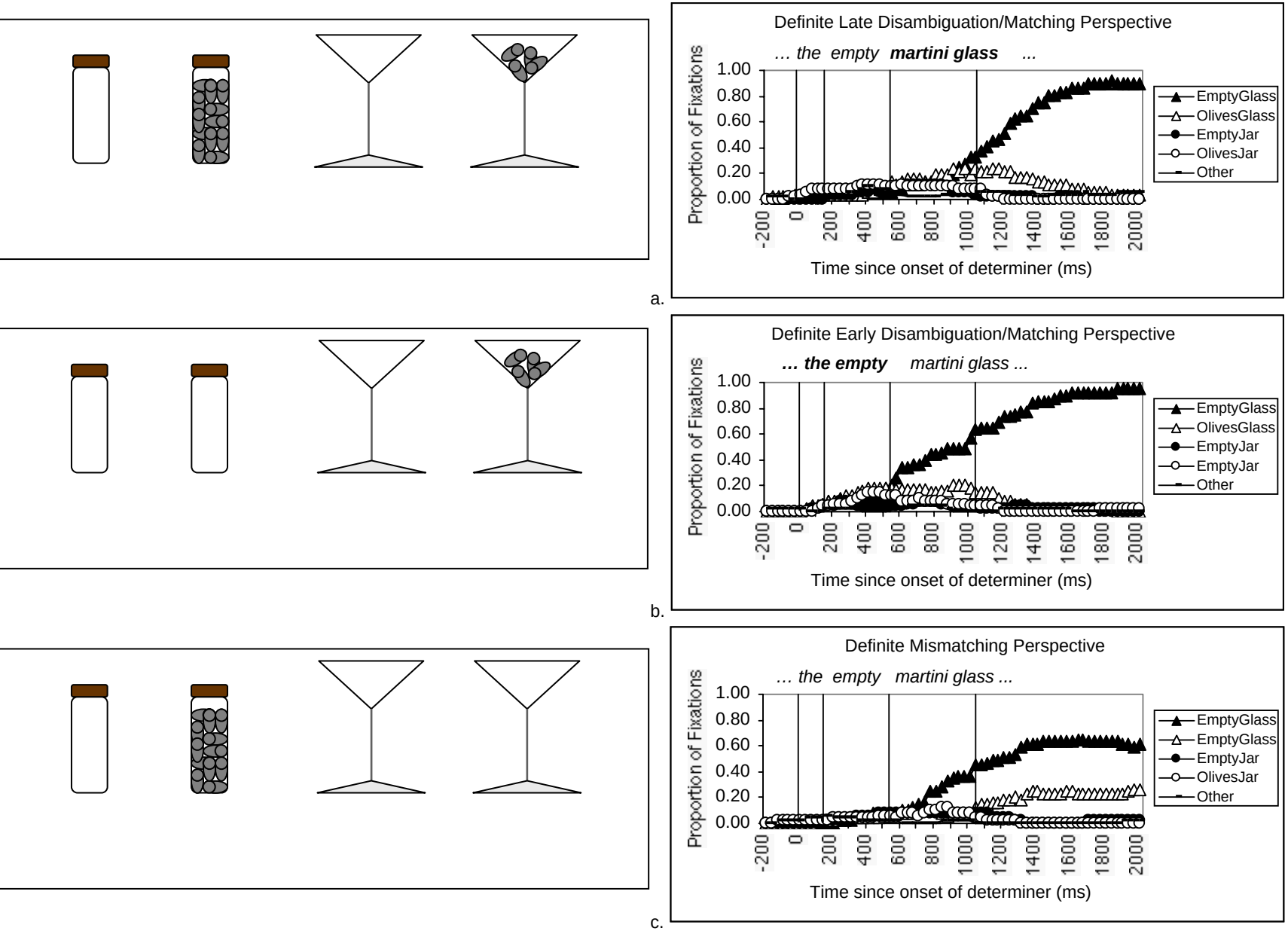


Figure 9. Left panels show sample displays for the definite late disambiguation/matching perspective, early disambiguation/matching perspective, and mismatching perspective conditions. Right panels show proportions of fixations on each object type over time. Points of disambiguation in the instructions are indicated in bold.

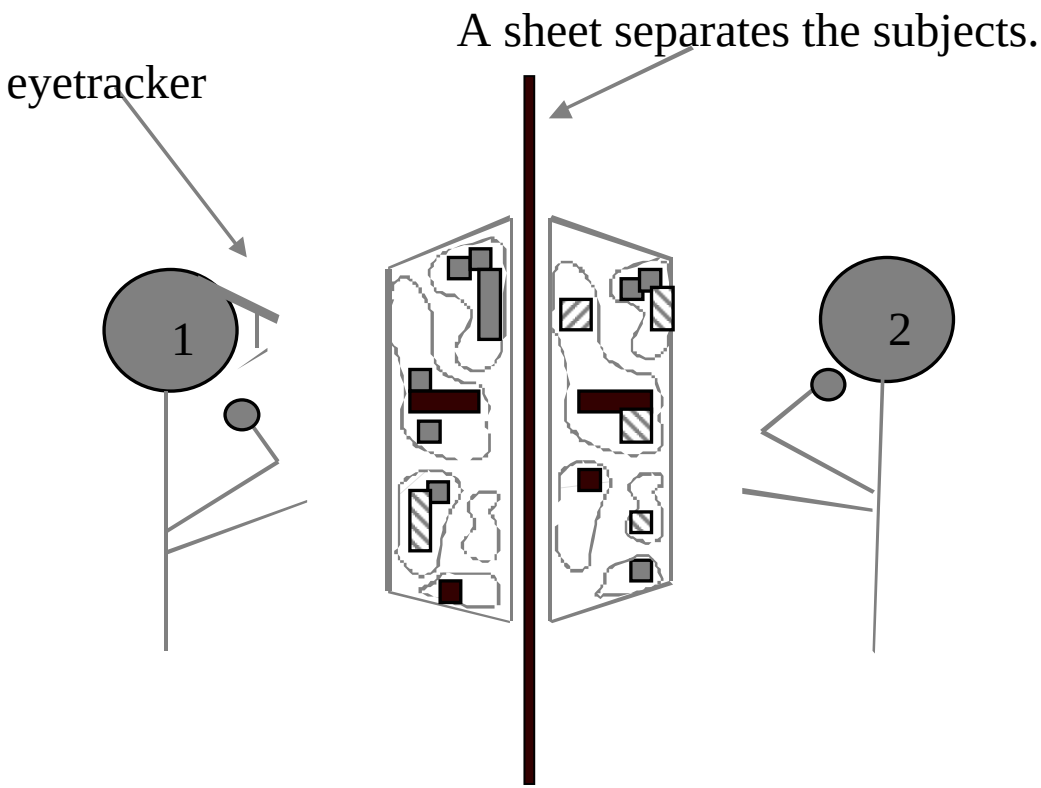
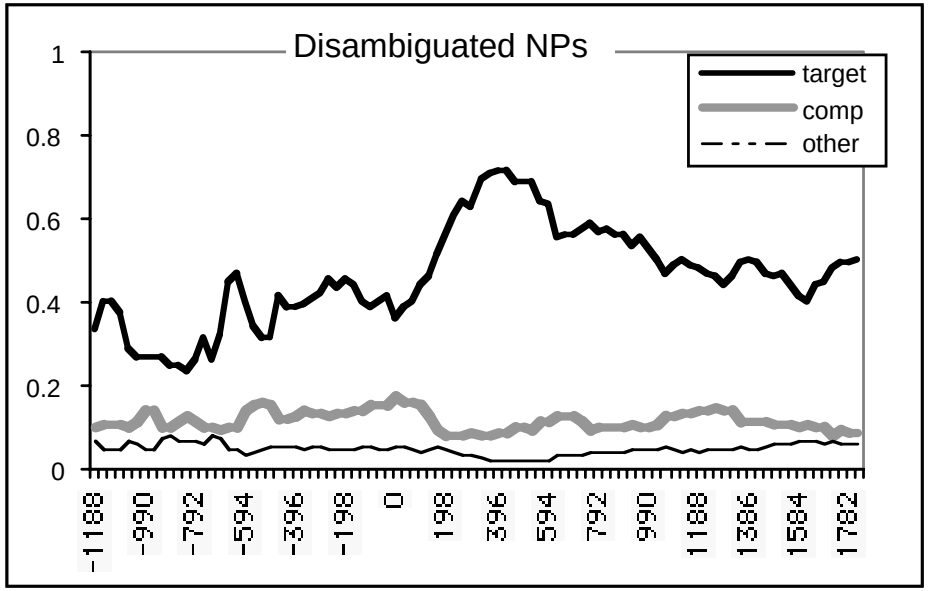


Figure 10. Schematic of the setup for the referential communication task. Solid regions represent blocks; striped regions represent stickers (which will eventually be replaced with blocks). The scene pictured is midway through the task, so some portions of the partners' boards match, while other regions are not completed yet.



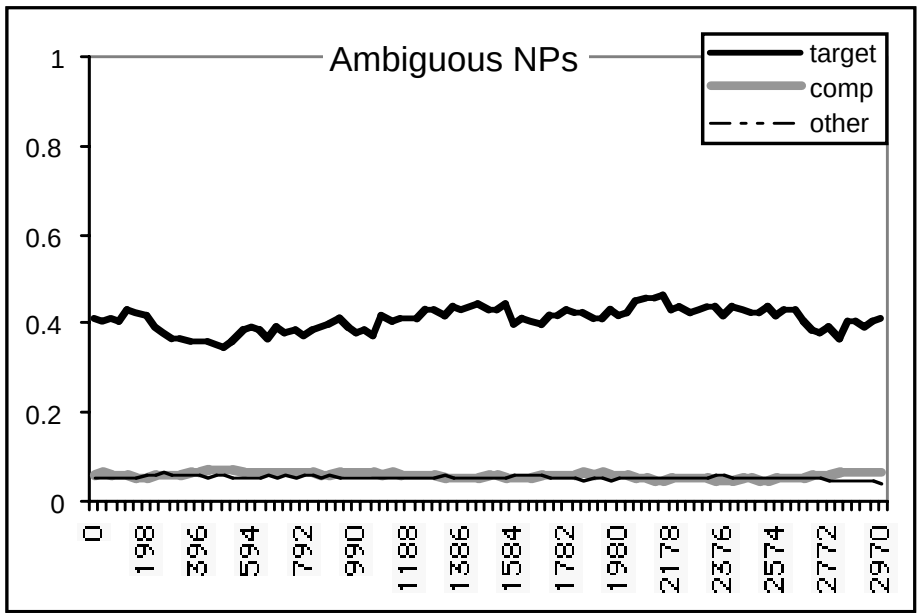


Figure 11. The top graph shows the proportion of fixations to targets, competitors, and other blocks by time (ms) for linguistically disambiguated definite noun phrases. The graph is centered by item with 0 ms = POD onset. The bottom graph shows the proportion of fixations for the linguistically ambiguous definite noun phrases.

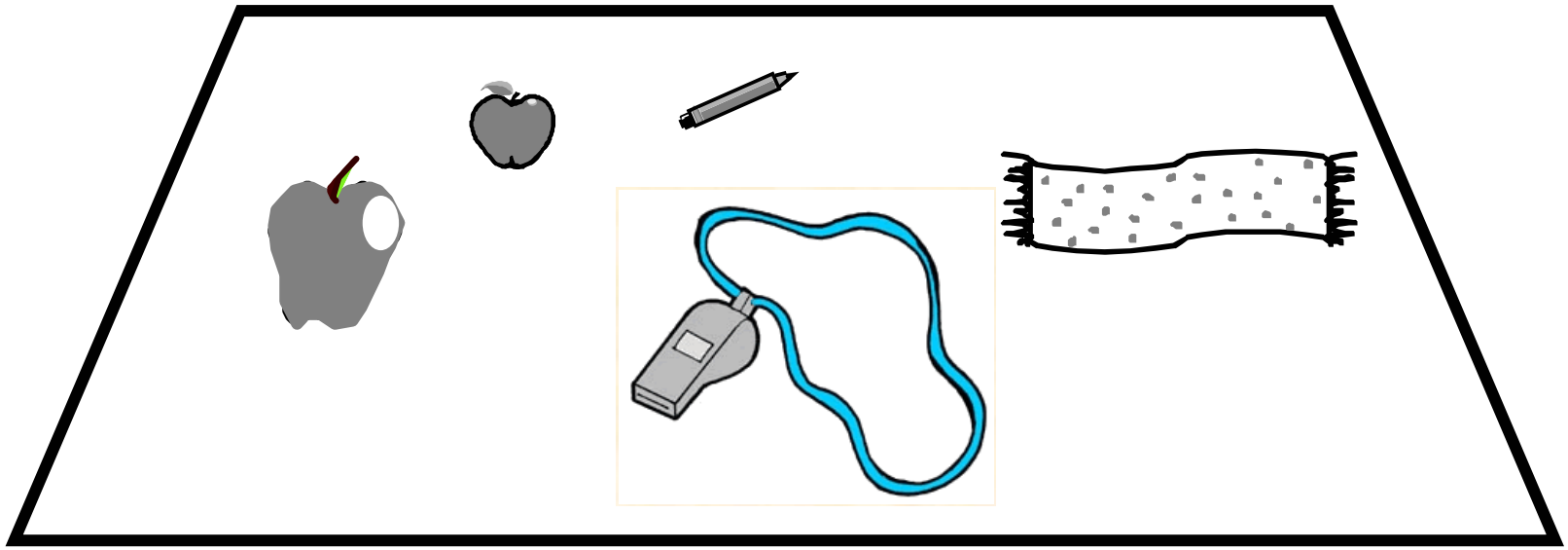


Figure 12. Hypothetical context for utterance *After John put the pencil below the big apple, he put the apple on top of the towel* to illustrate the implausibility of standard assumptions about context-independent comprehension. Note that the small (red) apple is intended to be a more prototypical apple than the large (green) apple.