



Published in final edited form as:

Lang Learn Dev. 2009 ; 5(4): 203–234. doi:10.1080/15475440903167528.

Use of Speaker's Gaze and Syntax in Verb Learning

Rebecca Nappa, Allison Wessel, Katherine L. McEldoon, Lila R. Gleitman, and John C. Trueswell

University of Pennsylvania

Abstract

Speaker eye gaze and gesture are known to help child and adult listeners establish communicative alignment and learn object labels. Here we consider how learners use these cues, along with linguistic information, to acquire abstract relational verbs. Test items were perspective verb pairs (e.g., *chase/flee*, *win/lose*), which pose a special problem for observational accounts of word learning because their situational contexts overlap very closely; the learner must infer the speaker's chosen perspective on the event. Two cues to the speaker's perspective on a depicted event were compared and combined: (a) the speaker's eye gaze to an event participant (e.g., looking at the Chaser vs. looking at the Flee-er) and (b) the speaker's linguistic choice of which event participant occupies Subject position in his utterance. Participants (3-, 4-, and 5-year-olds) were eye-tracked as they watched a series of videos of a man describing drawings of perspective events (e.g., a rabbit chasing an elephant). The speaker looked at one of the two characters and then uttered either an utterance that was referentially uninformative (*He's mooping him*) or informative (*The rabbit's mooping the elephant/The elephant's mooping the rabbit*) because of the syntactic positioning of the nouns. Eye-tracking results showed that all participants regardless of age followed the speaker's gaze in both uninformative and informative contexts. However, verb-meaning choices were responsive to speaker's gaze direction only in the linguistically uninformative condition. In the presence of a linguistically informative context, effects of speaker gaze on meaning were minimal for the youngest children to nonexistent for the older populations. Thus children, like adults, can use multiple cues to inform verb-meaning choice but rapidly learn that the syntactic positioning of referring expressions is an especially informative source of evidence for these decisions.

As a general consequence of the relational semantics they express, verbs serve as the linchpins for the combinatory meaning of sentences (e.g., Carlson & Tanenhaus, 1988; Grimshaw, 2005; Jackendoff, 2002). The successful comprehension of sentences typically hinges upon having detailed knowledge of a verb's syntactic and semantic preferences, such that both sorts of information are utilized to guide parsing and interpretation by adults (Altmann & Kamide, 1999; MacDonald, Pearlmutter, & Seidenberg, 1994; Tanenhaus & Trueswell, 1995) and by children (e.g., Trueswell, Sekerina, Hill, & Logrip, 1999; Snedeker & Trueswell, 2004).

In keeping with their eventual role in mature language use, the discovery procedures for verb meanings must coordinate information from the world to which the verbs refer, their distribution with respect to referring expressions, and their syntactic environments (Gleitman, 1990; Gleitman, Cassidy, Nappa, Papafragou, & Trueswell, 2005; Fisher, 1996). Moreover, other nonlinguistic sources of evidence might also be considered, including

possibly the perceived communicative intents of the speaker (e.g., Baldwin, 1991; Bloom, 2000). As we document experimentally in the present article, learners use these several sources of evidence selectively and in concert to solve the problems they face when learning lexical items, in large part because the gulf between these information types is often not as large or clear-cut as it first appears. We begin by discussing the word-learning problem as it is currently understood, focusing first on two general problems associated with the learning of any new lexical item. We then turn to how these problems manifest themselves when learning the meaning of a verb.

Two Problems in Word Learning

Two interlocking problems confront the learner upon hearing a novel word. The first, *the reference problem*, is to identify the thing (or action or relation or property) that the speaker is verbally describing. On the face of it this problem seems to be a difficult one even when the conversation is about the here-and-now, for under ordinary circumstances there are many objects and events within the listener's view, only one of which is the referent of the speaker's new word (cf. Locke, 1690/1964). The reference problem interacts with a second problem that is even more difficult. This is *the frame problem* (Fodor, 1987; Hayes, 1987) as applied to word learning. For even if the referred-to object or act can be selected accurately from the stream of objects and events, there are many ways to describe it; for example, every time an elephant is pointed to, a fortiori so is an animal and so is a mammal. Yet only one of these descriptions is intended by the speaker's choice of the word "elephant" (or "animal" or "mammal"). This choice cannot be apprehended (cf. Wittgenstein, 1953) or even differentiated (Quine, 1960) merely by noting what is being pointed at or looked at.

Both the reference problem and the frame problem have been intensely investigated experimentally over the past few decades, in an effort to understand how problems that philosophers have found so intractable for centuries are solved with such apparent ease by even the most average preschool-age children. One measure of children's success is that their rate of word learning is approximately eight words a day from the age of 2 years through, at least, early adulthood (Carey, 1982; Bloom, 2002). Most of the effort in understanding this remarkable feat has been directed toward the case of the concrete nouns that are the first lexical accomplishments of young children. For instance, approximately 75 to 80% of the first 100 words that Italian- and English-speaking children acquire are nouns, and less than 5% are verbs. This differs drastically from estimates of the linguistic input, in which parents' speech consists of approximately 30 to 40% nouns and 15 to 20% verbs (e.g., Caselli et al., 1995; Gentner & Boroditsky, 2001).

With respect to the reference problem as it influences the learning of nouns, several important studies have shown, first, that caretakers will often label a newly seen object using a characteristic linguistic format: "*Look! That's an elephant*" (e.g., Locke, 1690/1964; Quine, 1960; Tomasello & Farrar, 1986; Shipley & Shepperson, 1990). Infants and young children are closely attentive to several behavioral features that are associated with a speaker's attentive stance during these so-called ostensive labeling events. These include such social-attentive features as body orientation, pointing, and gaze directed toward the reference object (e.g., Baldwin, 1991; Bloom 2000; Bruner, 1983; Nurmsoo & Bloom, 2008; Tomasello, 1992).

As for the more troubling frame problem, several investigators have shown that there are strong representational biases that tend to mitigate situational ambiguity and narrow the descriptive choices of both child and adult learners. For example, learners preferentially interpret a new word as referring to the whole object in view at some middling level of abstraction (the basic-level category bias, cf. Rosch, Mervis, Gray, Johnson, & Boyes-

Braem, 1976), that is, viewing it as an elephant rather than, say, its superordinate, its parts, its color, or the substance of which it is made (Markman, 1990). Related examples include a preference for object names defined by shape rather than size or texture (Landau, Smith, & Jones, 1988) and animal names defined for the entire life cycle (*dog*) rather than a part of this cycle (*puppy*) (Hall & Waxman, 1993). Moreover, statistical schemas have been developed that try to account for how alternate descriptions can be eliminated in the course of several exposures to the word: Though one may initially hypothesize that ‘elephant’ is the interpretation of a word heard in the presence of that creature, subsequent hearings of the same word in the presence of an aardvark or a gnu can cause the learning machinery to back off to a more general reading, e.g., ‘animal’ (Harris, 1957; Siskind, 1996; Frank, Goodman, & Tenenbaum, 2007; Xu & Tenenbaum, 2007a, 2007b; Yu & Smith, 2007). Finally, once the learner has a vocabulary that includes members of more than one grammatical class, a considerable degree of linguistic cross-comparison among constituents is possible (cf. Harris, 1957). For example, the words *chef*, *bake*, *eat*, *cake* are co-predictive, as has been shown in both eye-tracking and priming studies (Altmann & Kamide, 1999; Ferretti, McRae, & Hatherell, 2001), allowing the learner to make intelligent inferences about a novel word in context.

Beyond Object Labels

Questions arise as to how far, and in what ways, the mechanisms and cues for word learning just mentioned scale up and generalize past the concrete basic-level whole-object terms that have been the focus of much of the early research on child word learning. The same problems clearly will arise in understanding the more abstract nominal vocabulary, including words such as *pet*, *home*, and *idea*.¹ In the present work, following very vigorous efforts in this regard (see particularly the collections in Tomasello & Merriman, 1995; Hirsh-Pasek & Golinkoff, 2006), we will be looking at aspects of verb learning. Though, as just mentioned, there is a relatively protracted period in the first half of the second year of life when verbs are rare in child speech, they begin to appear shortly thereafter. By the third birthday, the proportions of nouns and verbs in child speech roughly mirror the proportions in adult speech to children (Gentner & Boroditsky, 2001). For this later age then, we can now ask: Do the mechanisms and cues that subserve early noun learning play the same roles in verb learning?

Consider first the power of social-attentional cues for the acquisition of verbs, including the eye-gaze cue that has been found to be so potent in noun learning. A cautionary note is struck by the finding that caregivers do not ostensibly label and point to ongoing actions (“*Look! That’s eating!*”) with anything near the frequency, or in the close time-lock of word and gesture, with which this happens during object labeling (Gleitman, 1990; Tomasello & Kruger, 1992). Ostensive labeling of a co-present action is not, however, the only way that such social attentional cues can guide a listener/learner to the intended meaning of an unfamiliar verb. For instance, Akhtar and Tomasello (1997) found that children can use speaker eye-gaze cues indirectly to solve the reference problem for action words. In their study, children learned nonverbally to associate a novel event (e.g., launching a doll named Ernie into the air) with a novel instrument (a toy catapult). Later in the experiment, when a

¹In fact, it is remarkable how few words from any lexical class are free of the mind-driven complexities that even these quite simple examples exhibit. Consider a nominal expression such as “trash can” which describes not only the object but also its function in human affairs (if there were no people, the objects we call trash cans would surely continue to exist but they would no longer be trash cans). So too for the action verbs studied in these experiments, whose expressive content goes beyond the physical act (e.g., movement of an object between a source and a goal) to several mind-driven properties, of which the viewer-perspective property of give/get is only one. For instance, as we have also pointed out elsewhere, maybe you really can get blood from a stone but you can’t—grammatically speaking—give it any: The illformedness of “I gave some blood to the stone” demonstrates that the recipient of giving is necessarily a sentient being.

speaker uttered a sentence that used a novel verb for the first time (“*Let’s meek Ernie*”), the child was more likely to connect that verb to the previously observed novel action if the speaker had shifted her gaze to the event-relevant instrument (i.e., the catapult with which Ernie had previously been launched). These findings suggest that, although eye-gaze cues are neither as straightforward nor as frequent for verbs as they are for nouns, they do occur naturally and can be recruited by young children for verb-reference purposes.

How about the frame problem? This arises in the case of verbs as strongly and as variously as it does for nouns. For instance, consider an event in which George transfers possession of a ball to Anna. This is an event of *giving*, to be sure, but it is also necessarily an event in which the ball *moves* (a more general action term), and also an event in which either George *hands* or *throws* the ball to Anna (terms that express the manner of the motion) and *zips* or *lobs* the ball to Anna (expressing the rate or trajectory of motion) and an event in which Anna *receives* or *gets* the ball (terms that express the recipient’s accomplishment).

The example verb *give*, and related items that we next discuss, poses both a problem and an investigative opportunity for understanding verb learning that we will exploit in the experiments that follow (see Landau & Gleitman, 1985, and Gleitman, 1990, for earlier discussion of the logic of these predicate types). Because *giving* is a physical activity predicate, it occurs in situations that appear to exemplify its interpretation in quite a straightforward way, as we just described. Thus, unlike many verbs whose situational concomitants are obscure (imagine—or try to imagine—the typical situations in which *seem* or *know* are uttered), there are rather systematic behavioral cues to the use of this verb. This is convenient for examining effects of mutual attention, such as eye gaze, in supporting referential accuracy. But at the same time *give* poses a singularly intractable frame problem because it has a twinned alternate that occurs under highly overlapping, if not identical, circumstances, namely, *get* (and *receive*). Consider again George and Anna’s transfer of a ball. *Give* describes this event from the point of view of George: He is the giver, the grammatical Subject of the giving event. *Get* describes the same transfer of the same object, but taking the perspective of Anna: She is the getter/receiver, the grammatical Subject of the getting event. So for this pair of words, the referred-to event is identical but there is a difference in its framing: the perspective of the speaker. Because that perspective is not intrinsic to the event itself, but is rather in the mind of the speaker, how is the listener to perceive this distinction and acquire the word? That is: How is the toddler to realize that the English sound segment /gIv/ is the one that means ‘give’ whereas the sound segment /gEt/ is the one that means ‘get’? The only way is by becoming a mind reader and divining the speaker’s communicative intentions. We now briefly discuss the three possible ways (that we know of) by which listeners could come to read these intentions of speakers in the relevant regards so as to acquire the verb meanings.

Conceptual bias

Much as for the case of nouns, where powerful linguistic-conceptual biases such as the “whole object constraint” support early word learning, there are biases in understanding activities as well, and particularly for the perspective-verb pairs that we have been discussing. To see this point, consider Figure 1a which depicts another perspective event, a rabbit who is *chasing* an elephant—and in virtue of that, an elephant who is *fleeing* (from) a rabbit. Prior research (e.g., Fisher, Hall, Rakowitz, & Gleitman, 1994; Gleitman et al., 2005) has shown that an event depiction of this type, ostensibly labeled (“*This is meeking. What do you think meeking means?*”), is far more likely to elicit the conjecture *chase* in preference to *flee* or *run away*. Similarly, *give* is overwhelmingly preferred to *get* in similar, apparently ambiguous, visual circumstances.

The generalization that underlies these choices across the perspective verbs (*lead* versus *follow*, *win* versus *lose*, are some further instances) is quite subtle but has been investigated in detail by Lakusta and Landau (2005; see also Lakusta, Wagner, O’Hearn, & Landau, 2007). Their finding is that there is a strong preference for source-to-goal (to-path) or instigator-patient verbs such as *give* and *chase* over the reverse (from-path) descriptions in which the action is conceptually framed from the standpoint of the recipient (the getter, the flee-er). Indeed, such forward-looking verbs outnumber the backward-looking from-path verbs in the languages of the world; moreover, as again these investigators have shown, the argument encoding the source is much more likely to be omissible, preserving grammaticality, than is the argument encoding the goal. One might consider the idea that young learners acquire, for example, *give* before *get* and *chase* before *run away* because they discover (and track) the frequency differences for the two types in the exposure language, rather than having a representational bias to begin with. However, extensive study with infants secures that this asymmetry in the preferred encoding of sources and goals in locative events is present (and potent) prelinguistically (Csibra, Biro, Koos, & Gergely, 2003; Lakusta et al., 2007; Woodward, 1998, inter alia) and is continuous with the same bias documented in adults (e.g., Gleitman et al., op.cit.).

For the purpose of solving the mapping problem for verb learning, the preexisting to-path bias should trivialize the acquisition of *chase*, but the same bias should mislead the learner when the caretaker says *run away*. Because children by age 3 to 4 years are well in control of verbs of both of these kinds, and because confusions between them are rare and short-lived, it is clear that “something else” is aiding the learner in overcoming the conceptual bias so as to learn the dispreferred members of these perspective-verb pairs.

Attentive stance I: Eye gaze

In the case of perspective-predicate pairs, one such further possible source of evidence for solving the frame problem could be the visual-attentional stance adopted by the speaker. We know from recent work that adults (just prior to speaking) are much more likely to look at the sentential Subject of their sentence than at any other depicted character (Gleitman, January, Nappa, & Trueswell, 2007; Griffin & Bock, 2000). That is, they focus their attention on the entity they are talking “about” (see Talmy, 1978; Jackendoff, 2002; Gleitman, Gleitman, Miller, & Ostrin, 1996, for discussions of this Subject/Complement interpretive asymmetry). If, at least probabilistically, care-takers about to say *chase* look at chasers and caretakers about to say *flee* look at flee-ers, an attentive learner might, after several exposures, make just the right inference. Indeed, children are sensitive to eye gaze, head posture, and gesture when inferring a speaker’s referential intentions in object labeling (e.g., Baldwin, 1991) and interpreting actors’ intentions in actions (e.g., Carpenter, Akhtar, & Tomasello, 1998; Gergely, Bekkering, & Király, 2002; Woodward & Sommerville, 2000). This reading of actors’ intentions has been shown to narrow the scope of candidate concepts considered by the child for verb learning (e.g., Papafragou, Cassidy, & Gleitman, 2007; Poulin-Dubouis & Forbes, 2002; Tomasello & Barton, 1994). It is possible therefore that a speaker’s gaze direction toward a component of an event might be a giveaway for his or her communicative intent, in this case, the choice of event perspective.

Attentive stance II: The syntax-semantics mapping

As just discussed, sentence Subjects encode the “aboutness” of a predication and draw special attention. Once armed with the knowledge that English is a Subject-Verb-Object (SVO) language, the learner-listener inspecting the depiction in Figure 1 could therefore use the speaker’s choice of a sentential Subject to infer event perspective: If the Noun Phrase (NP) *the rabbit* is the Subject, the verb must mean ‘chase,’ and if instead *the elephant* is the Subject, it must mean ‘run away’. Indeed, past work has shown that children as young as 3

years of age make just these inferences. Fisher et al. (1994) had children watch puppets performing *chase/flee* and other perspective-predicate actions, while another puppet described each ongoing action. When the puppet said, “*Look! Blicking!*” both children and adults were more likely to think “blick” meant ‘chase’, in accord with the source-to-goal bias documented by Lakusta and Landau (2005). This tendency was further enhanced when the introducing sentence was, “*The rabbit is blicking the elephant,*” for this sentence is congruent with the interpretive bias. But when the puppet said, “*The elephant is blicking the rabbit,*” children shifted preponderantly (and adults just about absolutely) to ‘run away’ interpretations. Fisher et al. (1994) called this effect the **syntactic zoom lens** because the choice of sentential Subject cued the element chosen as the conceptual figure in the framing of the event (Talmy, 1978), thus fixing the interpretation of the novel verb. In short, an overt choice of grammatical Subject serves to determine the speaker’s attentive stance (i.e., what he is “talking about”) in much the same way—though perhaps to a different degree—as eye-gaze direction.

We emphasize that this hypothesized syntactic zoom lens procedure is not a simple one in the sense of exploiting a single evidentiary source. For one thing, the words *rabbit* and *elephant* would appear in the speaker’s utterance whether the verb meant *chase* or *flee*. Moreover, the particular transitive syntax is the same for both of these verbs, so the syntax alone does not resolve the meaning interpretation any more than the selection of the two nouns does. No more could the lexically complete syntactic analysis (taking the noun meanings and the syntax together) reveal the choice between these meanings. Rather, this entire structure has to be mapped against the co-occurring event (as in Figure 1) to resolve the issue. “*The rabbit is blicking the elephant*” can mean ‘*The rabbit is chasing the elephant*’ if and only if the depiction is as in Figure 1a; but it means ‘*The rabbit is fleeing the elephant*’ if the scene is as in Figure 1b. Thus the problem is solvable only by coordinating information across two cue types, one situational and the other syntactic.

In the experiments next presented, we explore verb-learners’ use of the potential sources of information just discussed. In all cases, we measure the effects of attentive stance by their influence on the (prenormed) conceptual bias in children’s learning of novel perspective verbs. Experiment 1 looks at children’s sensitivity to the speaker-eye-gaze cue in the absence of differentiating linguistic evidence (i.e., when the noun phrases are pronominal and therefore do not pick out the event/entity playing the subject role). Experiment 2 fully crosses eye gaze with the presence/absence of informative evidence from the structural positioning of referring noun phrases.

A possible outcome is that, when these cues are pitted against each other, syntactic evidence will take a backseat to head/eye gaze information when informing verb choices by very young children. Such a pattern would be consistent with a strong version of the so-called social-pragmatic theory of language acquisition in which intention reading from the social and situational context bears the earliest and heaviest weight in child word learning (e.g., Akhtar & Martínez-Sussmann, 2007; Bruner, 1983; Clark, 1997; Tomasello, 1992, 2003). The analysis of the syntactic zoom lens procedure sketched earlier might appear suggestive in the same direction: Because the structure-aided machinery is so complex, one might venture as a first hypothesis that it does not come into play as part of the lexical learning procedure until the learner has advanced past some logically prior novice state. As we will show, however, the structure-semantic linking cues appear to be prepotent even for very young experimental participants identifying the members of perspective-verb pairs. In the discussion, we defend the view that the findings for these very special items throw new light on general procedures for lexical acquisition.

EXPERIMENT 1: SPEAKER GAZE

In this experiment, participants viewed a video of an actor describing a stationary drawing of a perspective event (e.g., a rabbit chasing an elephant, see Figure 1a). The actor's utterance was always transitive and contained undifferentiating pronouns as the nominals (e.g., "He's blicking him"). The transitive (two argument) structure reduces the likelihood of experimentally irrelevant conjectures such as "They're running," yet is neutral about the two possible perspective verbs (e.g., chase/runaway) that would fit the video. The pronominals do not distinguish between pertinent entities performing the conjectured action. All else being equal, children should exhibit source-to-goal conceptual biases that favor, for example, chasing over fleeing meanings of *blick*. Bias levels for each perspective verb pair were pre-normed in a separate experiment using adult participants only. A single nonlinguistic cue to the actor's event perspective was provided by having the actor look at either one or the other character before he uttered the sentence. The goal was to establish whether (and if so, at what age) children use the speaker gaze cue to enhance or override conceptual biases.

Method

Participants—Thirty-nine native-English-speaking children participated in this study and provided data for the analyses below: eleven 3-year-olds (7 male, mean age 3;6), twelve 4-year-olds (9 male, mean age 4;5), and sixteen 5-year-olds (5 male, mean age 5;9). All were being raised as native speakers of English.

Data were collected at daycares and preschools in the Philadelphia area. Prior to participation, written parental consent was obtained using procedures approved by the University of Pennsylvania's Institutional Review Board.

Apparatus—A Tobii 1750 remote eye-tracking system was used. This system does not use an eye-tracking visor and instead tracks both eyes and head position via optics embedded in a flat panel computer display. Two laptop machines were used, both running the Windows XP operating system. One laptop displayed the video stimuli on the eye-tracker screen (via the Tobii Inc. Clearview software). A second laptop was dedicated to acquiring the eye-tracking data from the Tobii eye tracker (via the TET-server software, developed by Tobii, Inc.). The data sampling rate was 50 Hz, and spatial resolution of the tracker was approximately 0.5 degrees visual angle.

Procedure—Participants were run individually. Each child briefly played with the two experimenters and then was seated in a car booster seat placed in front of the Tobii 1750 eye-tracking monitor. Children were told they were going to watch television and play a fun game. The eye tracker was first calibrated by having the child watch a moving blue ball on the computer screen (i.e., a standard 5-point calibration procedure, devised by Tobii within their ClearView software). Viewing distance was approximately 24 to 36 inches.

The child then watched a prerecorded video involving two male actors, narrated by a female voice. The video depicted a story that was designed to explain the task to the child and motivate responsive behavior. First, an actor named Andy was introduced: "This is Andy. Andy likes to get things done." Then an actor named Magico was introduced: "This is Magico. Magico is a magician, but he isn't very good at magic." The narrator then explained that Magico likes to help Andy get things done but often causes unintended trouble. For instance, Magico is shown using magic to help Andy clean up a room, but instead causes a greater mess. Magico then tries to help Andy "work on his computer." Here Magico casts a spell which accidentally transports Andy inside the computer: Andy disappears (in a bright flash) from in front of his computer and then reappears on the computer screen, looking

quite surprised, peering out of the screen. This procedure was important for establishing the pretense that Andy's own perspective was from inside the computer screen, and he could observe other things that were also on the screen.

Specifically, Andy was then displayed above a clip art image (Figure 2). He looked down at the image (via a head turn and eye gaze shift) and said, "They're mooping." The female narrator then says, "Oh no! The spell has made Andy talk funny. The only way to break the spell and get Andy out of the computer is to figure out what Andy is saying! Can you tell us what mooping means?" If the child did not respond, the experimenters encouraged the child to respond and repeated the video. In practice, children readily understood (and enjoyed) the task. They typically responded with phrases such as "dancing" or "mooping means dancing."

After this introduction, the experiment proceeded with additional very similar trials of Andy describing a clip-art image. Each utterance contained a novel verb. A different novel verb was used on each trial, with one exception.² The child's spoken responses were recorded by hand and later coded for verb choice. At the end of the experiment, Andy was shown transported back out of the computer. The child then heard the narrator say, "Good job! You helped Andy get out of the computer!" The entire procedure lasted approximately 15 minutes.

Stimuli and Experimental Design—Target images consisted of perspective-predicate pairs. An example appears in Figure 1. These images had first been developed for an adult sentence production study (Gleitman et al., 2007). For the purposes of that study, the images were first normed by having a separate group of adults write down a single sentence description of each image. Items had to meet the following criteria to be selected for the present experiments: (a) For each image, the majority of written responses had to involve the expected perspective predicates (e.g., *A rabbit is chasing an elephant. An elephant is running away/fleeing from a rabbit.*). (b) Each image had a dominant verb response, henceforth, an A-response (e.g., *chasing*) and a subordinate verb response, henceforth, a B-response (e.g., *running from*). (c) Images had to have at least two attested B-responses to be used in the study. Henceforth, the character serving as the Subject of the A-response or B-response will be referred to as Character A (e.g., the rabbit) or Character B (e.g., the elephant), respectively.

Some new images had to be constructed for the present study to make them interpretable by children. For instance, a win/lose item showing a boxer winning a boxing match was not used in the present study. The new modified stimuli were also normed using a separate set of adults and chosen in the same manner.

A total of seven items met these criteria and were deemed child friendly (see Appendix). These clipart images were then embedded into digital videos, with Andy describing the image from above (e.g., Figure 3). Each video started with Andy looking toward the camera from his perch inside the computer. Approximately 1 second later, he moved his head and direction of gaze downward toward one of the two characters, as the figure shows. Finally, after an additional second, Andy uttered a perspective-ambiguous sentence using pronouns and a novel verb (e.g., *He's blicking him.*) Andy's shift in gaze was aligned with characters in the image by providing the actor with an off-camera object to fixate, positioned in the exact location of the intended referent within each image.

²Two target trials depicted chasing (one a dog chasing a man, another with a rabbit chasing an elephant). These two trials used the same novel verb (blick). One trial always appeared early in the experiment, the other late in the experiment. Informal inspection of the results from these trials showed no noticeable deviations from the other items.

Four presentation lists were constructed. For List 1, four of the test trials showed Andy looking at Character A, and the rest showed him looking at Character B. The target trials were placed in a fixed random order, embedded among one practice trial and six filler trials. Filler trials involved nonperspective-predicate items, also described by Andy using novel verbs and pronouns. List 2 was identical to List 1, except that Andy's direction of gaze was swapped on target trials (A-look videos were now B-look videos and vice versa). Two additional lists (Lists 3 and 4) were created by reversing the order of trials, to control for trial order. Each participant was assigned to one of the four lists.

Analysis

Eyetracking analysis: Each child's eye-tracking data were time-locked to the onset of each video. A time window of analyses was established that extended 4 seconds into each video. In this time window, track loss was determined separately for each eye by the ClearView Software. If track loss occurred for only one eye, the gaze coordinates from the other eye were used in analyses; if neither eye had track loss, the gaze coordinates from each eye were averaged. A given trial was dropped from further analysis if it had more than 33% track loss. If this process resulted in fewer than two trials within a condition for a given participant, the data from that entire participant were dropped; 3 participants had to be dropped under this criterion. For the remaining participants, an average of 8.3% of trials were excluded.

For each target video, three spatial scoring regions were defined: (a) Andy, which was the rectangular region toward the top of the screen within which he appeared; (b) Character A, which was an invisible rectangular region surrounding the character defined as the preferred sentential Subject in the preexperiment norming (see above); (c) Character B, which was an invisible rectangular region surrounding the character defined as the dispreferred sentential Subject in the norming. In a few instances, the shape of the character was asymmetrical and, as such, a single rectangular region could not surround it without including considerable white space. In these cases, two smaller, adjacent rectangular regions were used to tightly define a region around the character. Scoring regions were identical across the two experimental conditions within each item (that is, for example, the *chase* condition and the *flee* condition). Scoring regions were typically on the order of 2 to 3 degrees of visual angle in width and/or height.

Spoken response analysis: The verb meanings offered by participants were coded for each item as an A-response, a B-response, or NA. A-responses were verbs consistent with Character A being the sentential Subject of the event denoted by the verb in active form. B-responses were verbs consistent with Character B being the sentential Subject of the event denoted by the verb in active form. If the child did not offer an informative response (e.g., "*Mooping means mooping.*"), the experimenters asked, "What does mooping mean?" If the response was still uninformative, the experimenters asked, "Who is mooping?" If the response was Character A, then it was coded as an A-response; if it was Character B, then it was a B-response. NA responses were typically vague responses such as "moop means playing" for a *chase/flee* image.

Results

Spoken Responses—Figure 4 presents by Age group the mean proportion of A-, B-, and NA-responses as a function of the speaker's (Andy's) direction of Gaze toward Character A or Character B. The most prominent pattern in this figure is that A-responses (e.g., *chase*) predominated over B-responses (e.g., *run away*) for all age groups, just as in the norming data previously collected for adults. This outcome confirms the "preferred" and "dispreferred" framing choice within each of the seven predicate pairs: Even though the

images are “in principle” ambiguous, the children had a bias toward one of these, the same (source-to-goal) bias that has been shown for adult English-speaking populations.³

Another pattern in the data, of primary interest in the present context, is that all age groups showed the expected effect of Andy’s direction of gaze. When Andy looked at Character B, children regardless of their age provided fewer A-responses and more B-responses than when Andy looked at Character A.

To assess the reliability of this pattern, subject and item means were computed on the proportion of A-responses, and entered into separate analyses of variance (ANOVAs) each having two factors: Age (3yo, 4yo, 5yo) and Andy’s Gaze (A-Look, B-Look). As can be seen in Table 1, there were reliable effects of both Andy’s Gaze and Age group on A-Look responses (fewer A-responses occurred when Andy looked to Character B than to A, and older children provided more A-responses). The effect of Andy’s Gaze did not interact with Age group, indicating that speaker gaze influenced all ages. Separate ANOVAs for each age group verified that this was in fact the case, though the effect was nonsignificant in the 3-year-olds, $F1(1,10) = 2.21, p = 0.17$; $F2(1,6) = 3.72, p = 0.10$. To make sure that ANOVA effects were not distorted because of the use of proportions, we also computed an empirical logit transformation of the A-response data for each participant and entered the values into a weighted, repeated-measures generalized linear model (GLM) (see Barr, 2008).⁴ Consistent with the ANOVA results, logit A-responses were reliably predicted by Andy’s Gaze, Wald $\chi^2(1) = 17.05, p < 0.001$, and by Age Group, Wald $\chi^2(2) = 13.00, p < 0.005$, but not by the interaction term, Wald $\chi^2(2) = 1.70, p = 0.43$.

Eye Movements—Figure 5 presents the proportion of looks to Character A and to Character B over time relative to the onset of each video. (The average onset of Andy’s utterance from video onset was 84 samples [1680 ms]. Andy’s shift in gaze occurred approximately 1 second [50 samples] before that.) This graph collapses across all age groups to show the general pattern. As can be seen, children show an initial bias to look toward Character B (the goal of the action) followed by prolonged looks at Character A (the source and usually cause of the action). This pattern of early looks at the goal/patient/theme followed by looks to the source/agent has been seen in past studies with adults, when the task is not a language-production task but rather free viewing (Griffin & Bock, 2000; Papafragou, Halpern, & Trueswell, 2008). Of note, the strongest effects of Andy’s direction of gaze occurred only after Andy began speaking (rather than immediately after his shift of gaze). Specifically, after utterance onset looks to Character B increased when Andy was also looking at Character B as compared to when he was looking at Character A.

To assess the reliability of this gaze pattern, we computed subject and item means for the proportion of looks to Character B within three 500 ms time windows, relative to Andy’s utterance onset: –500 to 0 ms; 0 to 500 ms; and 500 to 1,000 ms.⁵ The subject mean proportions appear in Figure 6, broken down by Age, and the ANOVA results appear in Table 2.

³This bias held for all but one of the scenes (item 3 in the Appendix, *buy/sell*). For this item, the percentage of *buy* and *sell* guesses was approximately equal (about 40% each). We suspect this discrepancy has more to do with the particulars of the image we created, rather than the meanings of buy and sell, as a second buy/sell image (item 4 in the Appendix) had a large source-to-goal bias.

⁴This analysis was done in SPSS (v. 15.0) using the generalized estimating equations (GEE) module on subject means. An unstructured working correlation matrix was used. Other working correlation matrices generated significance patterns similar to those reported here.

⁵ANOVAs were performed for more traditional 200 ms time windows but did not reveal anything that the larger windows do not also show, so the analyses from the larger time windows are presented here. The general pattern is that participants look to Andy at video onset (they are cued with a crosshair to do so, so this is expected on multiple levels), then generally inspect the scene until he starts speaking, at which point they look back at him (particularly when he has used a pronoun), and follow his gaze and/or his linguistic cues to reference.

The pattern of means suggests that by the second half of the first second after utterance onset, all age groups are following the speaker's gaze (looking to B more when Andy is looking at B as compared to when Andy is looking at A). ANOVAs in the third time window (500–1,000 ms) confirmed the reliability of this effect. In the previous time window (0–500 ms), there was a reliable interaction between Andy's Gaze and Age, which reflects the fact that 4yo were, for some reason, slightly faster at following the speaker's gaze than the other age groups. To make sure that ANOVA effects were not distorted because of the use of proportions, we again computed an empirical logit transformation of the B-look data for each participant and entered the values into a weighted, repeated measures generalized linear model. Consistent with the ANOVAs, the first time window had no significant effects, the second window had a marginal effect of the interaction term, Wald $\chi^2(2) = 5.23$, $p = 0.07$, and the final region had a reliable effect only of the Andy Gaze predictor, Wald $\chi^2(1) = 13.57$, $p < 0.001$.

Finally, it should be noted that the proportion of looks to Characters A and B (in Figure 5) began near zero because participants were usually looking at the speaker (Andy) at the onset of the video. (Children were looking at Andy at the onset of approximately 55% of all trials, not plotted.) Early looks to the speaker are expected because the fixation cross prior to video onset was always located where Andy would appear. Participants quickly looked to the characters (as shown in Figure 4), and typically returned to look at Andy again after utterance onset, with the percentage of looks to Andy peaking at approximately 30%.

Discussion

When there was little in the linguistic signal to convey the speaker's perspective on an event, children (in all three age groups) relied on two factors to infer speaker perspective and verb meaning. First, they were heavily influenced by a conceptual bias toward source-to-goal predication. Second, they also used the direction of speaker's gaze to either strengthen (when Andy looked at speaker A) or mitigate (Andy looked at speaker B) this bias (Figure 4). Consistent with this, children's eye-movement patterns showed greater visual inspection of the character that Andy was also looking at, indicating a rapid alignment of perspective between speaker and listener, presumably based on physical cues such as speaker's head turn and gaze.

For the present case of verb learning, the findings indicate that children use speaker gaze information to do more than just establish object referents. Speaker gaze can affect what a listener thinks the sentence is about (and hence the referent that is intended to be in subject position). This being said, it is important to realize that these effects are quite weak, even in younger children who might be more likely to use a simple gaze heuristic. Thus an important concern that remains to be addressed is the potency of the eye-gaze cue in accomplishing the distinctions between members of perspective-verb pairs. After all, Figure 4 revealed that conceptual dominance (chase > flee) is still the major determinant of meaning conjecture even in the presence of countervailing eye-gaze information. Yet reports of confusion or mislearning within these pairs are rare; in the natural learning environment, children seldom use *give* when they mean something like 'get' or *win* when they mean 'lose'. To understand why these terms are so seldom confused, we next examine an additional factor that can determine the choice, namely, linguistic evidence about which character is the Subject.

EXPERIMENT 2: SPEAKER GAZE AND LINGUISTIC CUES

Experiment 2 mirrored Experiment 1, except that children now heard utterances overtly naming the characters the actor was viewing. In one condition, Character A was the sentential Subject (*The rabbit is blinking the elephant.*), and in the other condition Character

B was the Subject (*The elephant is blinking the rabbit.*). This factor was fully crossed in a 2 × 2 design with speaker gaze (A-Look vs. B-Look).

Experiment 1 established that children can use the social-pragmatic cue of eye gaze to infer verb meaning from perspective events, and past work has shown that children in this same age range can also use linguistic evidence to perspective taking (Fisher et al., 1994). The question explored here is how children weigh the social-pragmatic and linguistic evidence when both are present. Social-pragmatic approaches to verb learning (e.g., Bruner, 1983; Tomasello, 1992) would expect eye gaze to dominate linguistic evidence, especially in the youngest children, whereas theories that emphasize the informativeness of linguistic cues (syntactic bootstrapping, Gleitman, 1990; Gleitman et al., 2005) expect robust effects of the nominal cue to Subjecthood even in the youngest age group.

Method

Participants—Fifty-three native-English-speaking children participated: Twelve 3-year-olds (7 male, mean age 3;7), twenty-five 4-year-olds (19 male, mean age 4;5), and sixteen 5-year-olds (8 male, mean age 5;5).

Procedure—The procedure was the same as for Experiment 1.

Stimuli and Experiment Design—The stimuli were the same as Experiment 1 except for the following changes. Andy was videotaped describing the pictures using full noun phrases, rather than pronouns. For instance, in the warm-up video Andy said, “The man and the woman are mooping,” rather than, “They’re mooping.”

For target trials, four different videos were prepared. In the first, Andy looked at Character A and uttered a sentence with Character A in Subject position (A-Look with A-Syntax, e.g., he looked to the rabbit and said, *The rabbit’s mooping the elephant!*). The second version was the same as the first except Andy looked to Character B (B-Look with A-Syntax, e.g., he looked to the elephant and said, *The rabbit’s mooping the elephant!*).⁶ The third video was the same as the first, except that the order of the NPs changed, such that Character B was the Subject (A-Look with B-Syntax, e.g., he looked to the rabbit and said, *The elephant’s mooping the rabbit!*). The fourth video was like the third except that Andy looked to Character B (B-Look with B-Syntax, e.g., he looked to the elephant and said, *The elephant’s mooping the rabbit!*).

Experiment 2 had a total of eight stimulus lists: four stimulus organization lists were created, such that Andy’s direction of gaze was manipulated between subjects and Andy’s syntax was manipulated within subjects. In List 1, only A-Look target videos were used, four with A-Syntax, the rest with B-Syntax. These were randomly intermixed with the same filler trials used in Experiment 1. List 2 was the same as List 1, except only B-Look target videos were used. Lists 3 and 4 were the same as Lists 1 and 2, except that videos with A-Syntax were now used in B-Syntax and vice versa. The same order was used across all four lists. Each of these four stimulus organization lists was reversed to create four additional reverse-order lists, resulting in a total of eight lists, to which participants were pseudo-randomly assigned.

⁶Note that in traditional terms the syntax of *chase* and *flee* is always exactly the same (i.e., NVN), as is the mapping of this ordering onto argument structure (SVO). Only their functional structure (i.e., their semantic role assignments) differs, and accordingly, so do their NPs’ roles in the event as indexed in “the world”. We call this difference the “syntactic condition” as a shorthand. This is because the syntactic position of the relevant NP as sentential Subject (a) determines which of them the sentence is about, and, taken together with their indexing to the scene in view (b) fixes the verb as meaning either ‘chase’ or ‘flee’.

Analyses—Analysis procedures for both eye-tracking and spoken-response data in Experiment 2 were the same as those used in Experiment 1, as described above. Four participants were dropped from further analysis because of excessive trackloss (as defined in Experiment 1). For the remaining participants, an average of 6.4% of trials was excluded.

Results

Spoken Responses—Figure 7 presents by age group the mean proportion of A-, B-, and NA-responses as a function of the speaker's (Andy's) direction of gaze toward Character A or Character B (collapsing across type of syntax).

As can be seen in the figure, a very different pattern is observed from that obtained in Experiment 1. Here, when syntactic evidence is present, effects of Andy's Gaze on verb meaning choice are minimal to nonexistent. Younger children are the only ones showing any difference between Andy looking at Character A versus B.

For comparison, Figure 8 presents by age group the mean proportion of A-, B-, and NA-responses as a function of the speaker's (Andy's) syntax, A-Syntax or B-Syntax (collapsing now across eye gaze). Here we see very large effects of Syntax on verb meaning choice in all age groups. A-Syntax generates more A-responses, fewer B-responses, and fewer NA-responses as compared to B-Syntax.

Finally, Figure 9 presents by Age Group the full division of the data into the four conditions. Here, the most striking trends are the consistent effects of speaker syntax and the minimal effects of speaker eye gaze. It is only within the A-Syntax condition that we see some trends consistent with the use of eye gaze as a cue to verb meaning. In this condition, A-Look items generate more A-responses and fewer B-responses as compared with B-Look items.

Inferential statistical analyses of these data all point to large effects of Syntax and very little effect of Gaze on children's verb choices, regardless of age. Subject and item means were computed on the proportion of A-responses, and entered into separate ANOVAs having three factors: Age (3yo, 4yo, 5yo), Andy's Gaze (A-Look, B-Look), and Andy's Syntax (A-Syntax, B-Syntax) (see Table 1). Notably, there was a main effect of Andy's Syntax that did not interact with Age Group. This analysis also revealed no main effect of Age, no effect of Andy's Gaze, and no interaction between Age and Andy's Gaze. As can be seen in Figure 9, Andy's Gaze appears to have a small effect but only when he utters A-Syntax. This resulted in a reliable interaction between Gaze and Syntax in the item analysis but not the subject analysis (see Table 1). Such a pattern implies that subjects differed in whether they were sensitive to eye gaze in the presence of A-Syntax, but that Age group could not explain these individual differences (i.e., there were no interactions with Age, see Table 1). Thus, overall, children regardless of age respond robustly to the linguistic cues to verb meaning, and the presence of such evidence minimizes if not completely eliminates the use of eye gaze as a cue to verb meaning.

Like the ANOVAs, empirical logit A-responses were reliably predicted by Andy's Syntax, Wald $\chi^2(1) = 32.54, p < 0.001$, but not by Andy's Gaze, Wald $\chi^2(1) = 2.22, p = 0.14$, nor by Age Group, Wald $\chi^2(2) = 0.84, p = 0.66$, nor by any interaction terms (lowest $p = 0.14$). Separate analyses by age group (using subject and item ANOVAs and empirical logit GLM) showed values were reliably predicted by Syntax within each age group (all $ps < 0.05$), and by Gaze only in 4yo (all $ps < 0.05$, except the item ANOVA, in which the effect was marginally significant, $p = 0.10$). No age group showed an interaction between Gaze and Syntax.

In sum, when verb-relevant syntactic information was provided in the linguistic input (i.e., overt naming of which NP was the grammatical Subject and which the Complement), children did not use speaker eye gaze as a cue to verb meaning and instead used the syntactic-positioning cue almost exclusively. Trends in the means within each age group suggest that both 3- and 4-year-olds show a small susceptibility to the gaze cue. However, even in these younger age groups this trend is largely swamped by the distinct effect of syntactic cues (ANOVAs showed no interaction between Age and Syntax), indicating that regardless of age children are largely ignoring the gaze cue when interpreting the verb in the presence of informative linguistic cues.

Eye Movements—Figure 10 presents the proportion of looks to Character A and to Character B for all children (collapsing across age) as a function of time from video onset. This first figure plots only the main effect of Andy's Gaze (A-Look vs. B-Look) *collapsing across the two syntax conditions*. Somewhat surprisingly, children's gaze patterns (as a function of Andy's Gaze) are quite similar to what was seen in Experiment 1: effects of Andy's direction of gaze are seen after Andy began speaking. Approximately 15 samples (300 ms) after utterance onset, looks to Character B increased when Andy was also looking at Character B. Thus, even though children are not using speaker gaze to inform verb choice, they are nevertheless following speaker gaze with their eyes (looking where Andy is looking), presumably as a first clue to reference making.

For comparison, Figure 11 shows the same time course data, except now split by Andy's Syntax (A-Syntax vs. B-Syntax), collapsing across Andy's Gaze. Here we see that Andy's Syntax is having a similarly strong effect on listeners' eye-gaze patterns: More looks to the B character occur when Andy has uttered a B-Syntax sentence as compared to an A-Syntax sentence.

Figure 12 shows the complete breakdown of the data into the four conditions. Figure 12A shows effects of Andy's Gaze within A-Syntax utterances, whereas Figure 12B shows effects of Andy's Gaze within B-Syntax utterances. As one can see in the figure, after utterance onset there are effects of Andy's Gaze in both syntax conditions; however, the difference seems greater in the B-Syntax condition, mirroring the trends in the response data above.

To assess the reliability of this pattern, we computed subject and item means of the proportion of looks to Character B within three 500 ms time windows relative to Andy's utterance onset: -500 to 0 ms; 0 to 500 ms; and 500 to 1,000 ms. The subject mean proportions, by age group, appear in Figure 13 below and the results of subject and item ANOVAs appear in Table 3.

Although aspects of the bar graph are complex, the results from the ANOVAs reveal a clear pattern. Prior to utterance onset (-500 to 0 ms), children's direction of gaze is unaffected by the experimental factors. However, at utterance onset (0 to 500 ms), there is a strong influence of the speaker's (Andy's) direction of Gaze on children's gaze, which is followed immediately (500 to 1,000ms) by an equally strong effect of Andy's Syntax. In the absence of syntactic evidence to verb meaning (Experiment 1), effects of Andy's Gaze actually occurred *later* than what was observed here (during the 500 to 1,000 ms window, see Table 3). It is important to note, however, that the early effect of Andy's Gaze in the present experiment was driven almost entirely by the cue-concordant conditions, where the Syntax and Gaze cues matched (see Figure 13), resulting in a reliable interaction between Syntax and Gaze (see Table 3). This suggests that the syntactic evidence was actually modulating effects of speaker gaze. The timing of these main effects and interactions is consistent with when the cues could be detected by the child; Andy's Gaze shift occurs approximately 1

second prior to utterance onset, whereas Andy's Syntactic choice does not become apparent until he begins uttering the Subject Noun, which can affect the listener's gaze patterns during the perception of the word itself.⁷

As in Experiment 1, we also examined looks to the speaker (Andy) and found the same general pattern: early looks to Andy at video onset, followed by a sharp drop in looks to him (as participants looked at the picture) followed by looks back to Andy just prior to and just after Andy began to speak. Looks to Andy are especially interesting in the present experiment as they relate to the response given by the child. Does overriding the syntactic evidence in favor of gaze evidence coincide with disproportionate looks to the Speaker? Figure 14 plots looks to Andy relative to the onset of the first noun (N1) for the cue-conflict conditions, comparing trials on which participants provided an A response versus a B response. Indeed, providing a B response in the face of A-syntax was accompanied by more looks to Andy before and after the onset of N1. Providing an A-response in the face of B-syntax showed only a small increase in looks to Andy, which occurred only after hearing N1. We did not see a similar relationship between participants' responses and looks to the speaker in Experiment 1. Rather there were simply more looks to Andy overall. This no doubt reflects the fact that the syntactic position of the two referents could not be gleaned from the utterance in Experiment 1 without the use of speaker gaze cues, lending further support to our argument that use of speaker gaze cues in the present experiment became less relevant in the face of informative linguistic evidence.

Discussion

The major finding of Experiment 2 is that situational cues to interpretation—here, eye-gaze cues to the speaker's communicative intentions, cf. Baldwin, 1991—while heavily exploited under the partial-information situation of Experiment 1, are systematically demoted in the presence of internal linguistic information as to the meaning of perspective verbs. This conclusion holds even though there is a nonsignificant developmental trend in verb meaning guesses suggesting that the younger learners (the 3yo and 4yo groups) continue to show some influence of eye gaze in the cue-conflict condition of this experiment.

GENERAL DISCUSSION

Every experimental probe in our studies showed powerful effects of bias in the way humans conceptualize (“frame”) relations and events, and express these in sentences. Talmy (1978) has termed these representational biases Figure-Ground effects because of their close resemblance to the visual-spatial ambiguity findings in which such biases were first documented (see also Rosch, 1978; Tversky, 1977; Gleitman et al., 1996, for general discussion) Here we studied perspective verb pairs as a case in point. Though events of ‘chasing’ or ‘giving’ are also typically events of ‘fleeing’ and ‘getting’, respectively, speakers (including the young children we studied) heavily prefer the former interpretations. Our basic aim in the studies was to understand the cues that learners use to overcome these biases, such that they can learn the meanings of the words that describe the disfavored perspectives.

Experiment 1 was designed to ask about the potency of speaker eye gaze as a cue to the way a speaker is framing the event. Several studies in the literature have shown that eye gaze is

⁷The weighted empirical-logit analyses revealed similar patterns, though more effects were significant. In the first time window, there were no significant effects or interactions. In the second time window, Andy's Gaze was reliable, Wald $\chi^2(1) = 8.30, p < 0.005$, but so was Andy's Syntax, Wald $\chi^2(1) = 4.82, p < 0.05$, which reliably interacted with each other, Wald $\chi^2(1) = 6.92, p < 0.01$, and each interacted with Age, Wald $\chi^2(2) = 5.92, p = 0.05$; Wald $\chi^2(2) = 7.74, p < 0.05$. In the final time window, there were effects of Gaze, Wald $\chi^2(1) = 7.67, p < 0.01$, and Syntax, Wald $\chi^2(1) = 5.99, p < 0.05$, plus Gaze interacted with Age, Wald $\chi^2(2) = 5.92, p = 0.05$.

correlated with framing: Speakers whose eye gaze is subliminally captured on a depicted event participant are more likely to choose that participant as the Figure (hence grammatical subject) of their utterance (Gleitman et al., 2005). And infant and toddler listeners evidently think that their mothers' gaze direction indicates the referent of a new word (Baldwin, 1991). It is plausible, then, to expect that a speaker's gaze direction in a nonsense-verb situation will cue the listener as to which of two depicted entities is the Figure/Subject when the utterance itself (*He's blicking him.*) does not provide the solution. Experiment 1 confirmed the hypothesis that children's inference about a verb's meaning under these conditions of partial information was sensitive to the speaker's gaze: A speaker who looked at the chaser increased children's 'chase' responses and decreased 'flee' responses, and the opposite effect was found when the speaker looked at the flee-er (see Figure 4 and Table 1).

This manifest sensitivity to the speaker gaze cue was almost completely eliminated in Experiment 2, when linguistic evidence about the speaker's perspective was made available by labeling the Subject/Figure (e.g., *The rabbit's blicking the elephant.*). Children as young as we were able to test (3yo) treated this language-internal information as decisive even in the experimental condition where eye gaze was giving a contrary cue: Eye-movement analyses showed that while children continued to track the speaker's gaze in the presence of syntactic information (see Figure 10 and Table 3), they demoted the information this cue might provide in the face of countervailing syntactic evidence (see Figure 9 and Table 1).

We did nevertheless observe a nonsignificant developmental trend in this cue conflict condition: Younger children appeared to be more tempted by the speaker gaze cue than the older children. Specifically, when syntactic and gaze evidence conflicted, older children were more likely than younger children to use syntactic evidence to guide verb decisions, with no such developmental change occurring in cue-concordant conditions (see Figure 9 and Table 1). A developmental pattern of increased weighting of internal linguistic cues over observational ones is expectable, on several grounds. First, as the learner moves from an initial state in which the learning machinery seems to be almost completely limited to the acquisition of basic-level nouns (Gentner & Boroditsky, 2001) to a stage where first verbs are acquired, the relevant information sources for learning change accordingly. As an important instance, ostensive labeling of visually co-present events (as opposed to visually co-present objects) has been conjectured to be exceedingly rare (Gleitman, 1990), an observation that has been confirmed experimentally (Tomasello & Kruger, 1992). This is because speakers tend to talk about upcoming or past events rather than simultaneously occurring events. To the extent this is so, speaker gaze patterns provide less stable and less compelling learning indications for verbs than they do for nouns. Perhaps most relevant in the present context, prior evidence shows that an increasing reliance on syntactic information over observational cues would also be an appropriate adjustment by the learning machinery, given the nature of early-acquired and later-acquired items within the verb class itself. Snedeker and Gleitman (2004; see also Gillette, Gleitman, Gleitman, & Lederer, 1999; Papafragou et al., 2007) have shown that information available from observation of extralinguistic context is most useful for acquiring the concrete, early-learned verbs (e.g., *run, hold, kiss*), whereas most of the information for acquiring abstract, later-learned, verbs resides in their syntactic environments (e.g., *think, know*). In these learning-informational regards, the perspective verbs occupy an intermediate position, presumably owing to the special framing problem they pose. We expect, then, that as the learning machinery advances, its weighting of observational evidence, including eye-gaze cues, diminishes, at least in situations where syntactic information is available.

One possible objection to this conclusion could be the extent to which our experimental eye-gaze shift was a natural one. In order to control for gaze patterns, we had an actor generate a simple gaze cue: He shifted his head and eye gaze toward a referent and then, approximately

1 second later, uttered a description while holding gaze on that referent. Although this shift in gaze mimics natural patterns observed prior to utterance onset, speakers also tend to make a second shift in gaze, immediately after utterance onset, toward the second referent (the Object) in the sentence (Griffin & Bock, 2000; Gleitman et al., 2007). Also, recent work by Csibra and Gergely (2006) suggests the importance of pedagogical cues (such as specific patterns of eye contact and prosody) in encouraging children to make the correct inferences about the goals of an actor/speaker. It may also matter that the present study manipulated eye gaze via video presentation rather than through face-to-face interaction with a live actor. It is important that future work examine these more specific cues in a range of linguistic contexts. Given the present findings though, we strongly suspect that even if the influence of eye gaze on verb choice is enhanced by this improvement in technique, syntactic evidence will continue to exert the stronger and more dominating influence.

Other important issues remain open as well. For instance, there exists a second class of deictic-perspective predicates that might yield effects of eye gaze larger than found here, e.g., locative-stative predicates are routinely uttered in ostensive settings (e.g., *Where's my ball? It is under the sofa; Pick me up!*). Thus, children hearing locative predicates with a novel preposition (e.g., *The ball is agorp the sofa.*) ought to show stronger effects of speaker gaze than they do in event labeling. Past research has shown that children can use the relevant syntactic evidence to infer the meaning of novel locative prepositions in this context (Landau & Stecker, 1990; Fisher, Klingler, & Song, 2006). So it would be especially interesting to see how these syntactic cues combine with speaker gaze cues.

Finally, it is crucial for our understanding of the word-learning process to better understand the detailed nature of the syntactic evidence that children use to identify the grammatical Subject. In principle, word order and various coordinating properties of morphology can be informative for such choices. But understanding learning realistically depends on knowing the sequence in which such cues come on line. For instance, as early as 1982, Slobin and Bever were able to show that procedures involving word order were usable by young learners long before they could exploit morphological cues to passivization (see also Bates & MacWhinney, 1989, and, e.g., von Berger, Wulfeck, Bates, & Fink, 1996). Thus the presence of cues in the input does not guarantee their uptake. Significant headway on this class of problems has been made in recent work (e.g., Scott & Fisher, 2009; Yuan & Fisher, 2009; Fisher & Song, 2006). Final analysis in the terms of the cues that our young participants exploited to make their syntactic inferences awaits the outcomes of such detailed research.

General Implications

Taken as a whole, our results suggest that children, even at the earlier stages of verb learning, are sensitive to the informativeness of various cues to a verb's meaning and take this informativeness into account when converging on the intended meaning of a novel verb. When linguistic evidence is not very informative (as in the situation of Experiment 1, using pronominals that did not distinguish between the referents), children could still use extralinguistic evidence, such as physical cues to a speaker's attentional stance, including head posture and eye gaze, to support their word-meaning conjectures. It is important to notice that, just because not all listening situations are maximally informative, eye gaze (and related social-pragmatic cues) are bound to play a heavy role in conversational understanding throughout life, to reduce the indeterminacy of pronominal reference (as in Experiment 1), among other reasons (see, e.g., Hanna & Brennan, 2007, for one such study with adults). Thus the findings of Experiment 2 should not be interpreted as superseding those of Experiment 1, but rather as showing the contrast in interpretive procedures that is consequent to the amount of information available in particular listening situations.

Experiment 2 showed that when linguistic evidence for a framing choice is present, participants of all ages preferentially use this information almost to the exclusion of gaze- and gesture-directional evidence. As we have just discussed, not only the presence/absence of pronouns but also the nature of the items to be learned is also likely to influence which sort of evidence is informative. The framing ambiguity rarely arises for verbs such as *hop* or *kiss*, where observation of the contingencies for the word's use—bolstered by the “basic level” bias in interpretation—transparently yields unique interpretations (cf. Carey, 1982). But where framing questions are prominent in the acquisition problem, such as for the perspective verbs studied here, learning occurs somewhat later in development and draws more heavily on syntactic-semantic mappings than on social-pragmatic cues to the speaker's intent.

The experimental findings reported here do not speak directly to the question of cause and effect in such a reweighting of cues to verb interpretation over developmental time. It is possible that the learning machinery adjusts its weighting of cues in response to the changing nature of its present task (i.e., the requirement to acquire more and more abstract vocabulary items whose situational concomitants are less straightforward). But it is just as possible that the changing nature of the learning machinery (its enhanced ability, with increasing language-specific knowledge, to recruit syntax-semantic links as cues to verb meaning) is what causes the abstract items to become more easily learnable. However, a look at collateral findings on sentence processing development is suggestive of the causal flow underlying these cue reweightings. For instance, it has been posited that among the evidentiary sources available to the child at a given point in language learning, constraints on sentence meaning will be weighted based on the child's current estimation of their reliability (Trueswell & Gleitman, 2004, 2007). That is, children use the same constraint-satisfaction process for sentence interpretation that has been documented again and again in adults (e.g., Bates and MacWhinney, 1987; Seidenberg & MacDonald, 1999; MacDonald et al., 1994; Trueswell & Tanenhaus, 1994). In this case, while syntactic structure is a robustly trustworthy cue to a verb's meaning (e.g. Gleitman, 1990), gaze allocation often functions as an automatic process, reflecting not just immediate referential selection of a sentential Subject but also many different dimensions of cognition (e.g., Rayner, 1998; Trueswell, 2008; Yantis & Jonides, 1990, 1996); the prudent language learner, while attending to both these important cues, would rely much more heavily on explicit syntactic role assignment than on a transient and ambiguous gaze cue. Across the current findings and past experimental work on ambiguity resolution, listeners appear to use multiple evidentiary sources, weighted by reliability, to interpret an utterance (e.g., Arnold, Brown-Schmidt, & Trueswell, 2007; Hurewitz, Brown-Schmidt, Thorpe, Trueswell, & Gleitman, 2001; Trueswell et al., 1999; Snedeker & Trueswell, 2004). Listeners of all ages consistently strive to extract a coherent analysis of the utterance that creates the fewest possible and least egregious violations of syntactic, semantic, and referential constraints.

Acknowledgments

This work was partially funded by a grant to LRG and JCT from the National Institutes of Health (1-R01-HD37507). Special thanks to Dr. Andy Connolly for spending so much time trapped inside a computer for this project.

REFERENCES

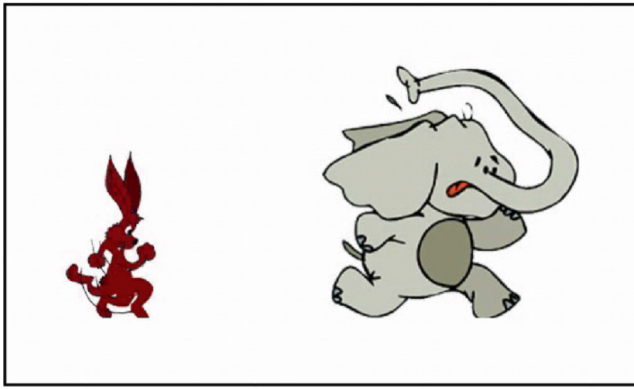
- Akhtar, N.; Martinez-Sussmann, C. Intentional communication. In: Brownell, CA.; Kopp, CB., editors. *Transitions in early socioemotional development: The toddler years*. New York: Guilford; 2007. p. 201-220.
- Akhtar N, Tomasello M. Young children's productivity with word order and verb morphology. *Developmental Psychology*. 1996; 33:952–965. [PubMed: 9383618]

- Altmann GTM, Kamide Y. Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*. 1999; 73:24.
- Arnold J, Brown-Schmidt S, Trueswell JC. Children's use of gender and order-of-mention during pronoun comprehension. *Language and Cognitive Processes*. 2007; 22(4):527–565.
- Baldwin DA. Infant contribution to the achievement of joint reference. *Child Development*. 1991; 62:875–890. [PubMed: 1756664]
- Barr D. Analyzing 'visual world' eyetracking data using multilevel logistic regression. *Journal of Memory and Language: Special Issue on Emerging Data Analysis and Inferential Techniques*. 2008; 59(4):457–474.
- Bates, E.; MacWhinney, B. Competition, variation, and language learning. In: MacWhinney, B., editor. *Mechanisms of language acquisition*. Hillsdale, NJ: Lawrence Erlbaum Associates; 1987. p. 157-193.
- Bates, E.; MacWhinney, B. Functionalism and the Competition Model. In: MacWhinney, B.; Bates, E., editors. *The crosslinguistic study of sentence processing*. New York: Cambridge University Press; 1989. p. 157-193.
- Bloom, P. How children learn the meanings of words. Cambridge, MA: MIT Press; 2000.
- Bloom P. Mindreading, communication, and the learning of the names for things. *Mind and Language*. 2002; 17:37–54.
- Bruner, J. *Child's talk*. New York: Norton Publishing; 1983.
- Carey, S. Semantic development: The state of the art. In: Wanner, E.; Gleitman, L., editors. *Language acquisition: The state of the art*. UK: Cambridge University Press; 1982. p. 347-389.
- Carlson, G.; Tanenhaus, MK. Thematic roles and language comprehension. In: Wilkins, W., editor. *Thematic relations*. New York: Academic Press; 1988. p. 263-288.
- Carpenter M, Akhtar N, Tomasello M. Fourteen- through 18-month-old infants differentially imitate intentional and accidental actions. *Infant Behavior and Development*. 1998; 21(2):315–330.
- Caselli MC, Bates E, Casadio P, Fenson J, Fenson L, Sanderl L, et al. A cross-linguistic study of early lexical development. *Cognitive Development*. 1995; 10:159–199.
- Clark EV. Conceptual perspective and lexical choice in acquisition. *Cognition*. 1997; 64(1):1–37. [PubMed: 9342930]
- Csibra G, Biro S, Koos O, Gergely G. One year old infants use teleological representations of actions productively. *Cognitive Science*. 2003; 27:111–133.
- Csibra, G.; Gergely, G. Social learning and social cognition: The case for pedagogy. In: Munakata, Y.; Johnson, MH., editors. *Processes of change in brain and cognitive development: Attention and performance XXI*. Oxford, UK: Oxford University Press; 2006. p. 249-274.
- Ferretti TR, McRae K, Hatherell A. Integrating verbs, situation schemas, and thematic role concepts. *Journal of Memory and Language*. 2001; 44:516–547.
- Fisher C. Structural limits on verb mapping: The role of analogy in children's interpretations of sentences. *Cognitive Psychology*. 1996; 31:41–81. [PubMed: 8812021]
- Fisher C, Hall DG, Rakowitz S, Gleitman L. When it is better to receive than to give: Syntactic and conceptual constraints on vocabulary growth. *Lingua*. 1994; 92:333–375.
- Fisher C, Klingler SL, Song H. What does syntax say about space? 26-month-olds use sentence structure in learning spatial terms. *Cognition*. 2006; 101:B19–B29. [PubMed: 16364280]
- Fisher, C.; Song, H. Who's the subject? Sentence structures as analogs of verb meaning. In: Hirsh-Pasek, K.; Golinkoff, RM., editors. *Action meets word: How children learn the meanings of verbs*. New York: Oxford University Press; 2006. p. 392-425.
- Fodor, JA. Modules, frames, fridgeons, sleeping dogs and the music of the spheres. In: Pylyshyn, Z., editor. *The robot's dilemma: The frame problem in artificial intelligence*. Norwood, NJ: Ablex; 1987. p. 139-149.
- Frank, MC.; Goodman, ND.; Tenenbaum, J. A Bayesian framework for cross-situational word learning. Paper presented at the 20th Advances in Neural Information Processing Systems Conference; British Columbia, Canada. 2007 Dec.

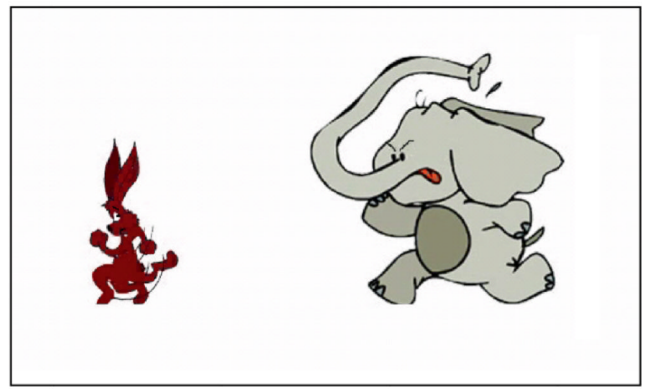
- Gentner, D.; Boroditsky, L. Individuation, relativity and early word learning. In: Bowerman, M.; Levinson, SC., editors. *Language acquisition and conceptual development*. New York: Cambridge University Press; 2001. p. 215-256.
- Gergely G, Bekkering H, Király I. Rational imitation in preverbal infants. *Nature*. 2002; 415:755. [PubMed: 11845198]
- Gillette J, Gleitman H, Gleitman LR, Lederer A. Human simulations of vocabulary learning. *Cognition*. 1999; 73:135–176. [PubMed: 10580161]
- Gleitman LR. The structural sources of verb meanings. *Language Acquisition*. 1990; 1:3–55.
- Gleitman L, Cassidy K, Nappa R, Papafragou A, Trueswell J. Hard words. *Language Learning and Development*. 2005; 1(1):23–64.
- Gleitman LR, Gleitman H, Miller C, Ostrin R. Similar, and similar concepts. *Cognition*. 1996; 58:321–376. [PubMed: 8871343]
- Gleitman LR, January D, Nappa R, Trueswell JC. On the give and take between event apprehension and utterance formulation. *Journal of Memory and Language*. 2007; 57(4):544–569. [PubMed: 18978929]
- Griffin ZM, Bock K. What the eyes say about speaking. *Psychological Science*. 2000; 11:274–279. [PubMed: 11273384]
- Grimshaw, J. *Words and structure*. Stanford, CA: CSLI Publications; 2005.
- Hall DG, Waxman SR. Assumptions about word meaning: Individuation and basic-level kinds. *Child Development*. 1993; 64:1550–1570.
- Hanna JE, Brennan SE. Speakers' eye gaze disambiguates referring expressions early during face-to-face conversation. *Journal of Memory and Language*. 2007; 57(4):596–615.
- Harris Z. Co-occurrence and transformation in linguistic structure. *Language*. 1957; 33:283–340.
- Hayes, PJ. What the frame problem is and isn't. In: Pylyshyn, Z., editor. *The robot's dilemma: The frame problem in artificial intelligence*. Norwood, NJ: Ablex; 1987. p. 123-137.
- Hirsh-Pasek, K.; Golinkoff, RM., editors. *Action meets word: How children learn verbs*. New York: Oxford Press; 2006.
- Hurewitz F, Brown-Schmidt S, Thorpe K, Gleitman LR, Trueswell JC. One frog, two frog, red frog, blue frog: Factors affecting children's syntactic choices in production and comprehension. *Journal of Psycholinguistic Research*. 2001; 29(6):597–626. [PubMed: 11196065]
- Jackendoff, R. *Foundations of language*. Oxford, UK: Oxford University Press; 2002.
- Lakusta L, Landau B. Starting at the end: The importance of goals in spatial language. *Cognition*. 2005; 96:1–33. [PubMed: 15833305]
- Lakusta L, Wagner L, O'Hearn K, Landau B. Conceptual foundations of spatial language: Evidence for a goal bias in infants. *Language Learning and Development*. 2007; 3(3):179–197.
- Landau, B.; Gleitman, LR. *Language and experience: Evidence from the blind child*. Cambridge, MA: Harvard University Press; 1985.
- Landau B, Smith LB, Jones S. The importance of shape in early lexical learning. *Cognitive Development*. 1988; 3:299–321.
- Landau B, Stecker D. Objects and places: Geometric and syntactic representations in early lexical learning. *Cognitive Development*. 1990; 5:287–312.
- Locke, J. *An essay concerning human understanding*. Cleveland, OH: Meridian Books; 1964. p. 259-298.(Original work published 1690)
- MacDonald MC, Pearlmuter NJ, Seidenberg MS. The lexical nature of syntactic ambiguity resolution. *Psychological Review*. 1994; 101:676–703. [PubMed: 7984711]
- Markman EM. Constraints children place on word meanings. *Cognitive Science*. 1990; 14(1):57–77.
- Nurmsoo E, Bloom P. Preschoolers' perspective-taking in word learning: Do they blindly follow eye gaze? *Psychological Science*. 2008; 19:211–215. [PubMed: 18315790]
- Papafragou A, Cassidy K, Gleitman L. When we think about thinking: The acquisition of belief verbs. *Cognition*. 2007; 105:125–165. [PubMed: 17094956]
- Papafragou A, Hulbert J, Trueswell JC. Does language guide event perception? Evidence from eye movements. *Cognition*. 2008; 108(1):155–184. [PubMed: 18395705]

- Poulin-Dubois P, Forbes J. Toddlers' attention to intentions-in-action in learning novel action words. *Developmental Psychology*. 2002; 38:104–114. [PubMed: 11806694]
- Quine, W. *Word and object*. New York: Wiley; 1960.
- Rayner K. Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*. 1998; 124:372–422. [PubMed: 9849112]
- Rosch, E. Principles of categorization. In: Rosch, E.; Lloyd, B., editors. *Cognition and categorization*. Hillsdale, NJ: Erlbaum; 1978. p. 27-48.
- Rosch E, Mervis CB, Gray W, Johnson D, Boyes-Braem P. Basic objects in natural categories. *Cognitive Psychology*. 1976; 8:382–439.
- Scott RM, Fisher C. 2-year-olds use distributional cues to interpret transitivity-alternating verbs. *Language and Cognitive Processes*. 2009; 24(6):777–803. [PubMed: 20046985]
- Seidenberg MS, MacDonald MC. A probabilistic constraints approach to language acquisition and processing. *Cognitive Science*. 1999; 23:569–588.
- Shipley ER, Shepperson B. Countable entities: Developmental changes. *Cognition*. 1990; 34:109–136. [PubMed: 2311354]
- Siskind J. A computational study of cross situational techniques for learning word-to-world mappings. *Cognition*. 1996; 61(1–2):39–91. [PubMed: 8990968]
- Slobin DI, Bever TG. Children use canonical sentence schemas: A crosslinguistic study of word order and inflections. *Cognition*. 1982; 12:229–265. [PubMed: 6891309]
- Snedeker, J.; Gleitman, L. Why it is hard to label our concepts. In: Hall, DG.; Waxman, SR., editors. *Weaving a lexicon*. Cambridge, MA: MIT Press; 2004. p. 255-293.
- Snedeker J, Trueswell JC. The developing constraints on parsing decisions: The role of lexical-biases and referential scenes in child and adult sentence processing. *Cognitive Psychology*. 2004; 49(3): 238–299. [PubMed: 15342261]
- Talmy, L. Figure and ground in complex sentences. In: Greenberg, JH.; Ferguson, CA.; Moravcsik, EA., editors. *Universals of human language (IV): Syntax*. Stanford, CA: Stanford University Press; 1978.
- Tanenhaus, MK.; Trueswell, JC. Sentence comprehension. In: Miller, JL.; Eimas, PD., editors. *Handbook in perception and cognition: Vol. 11. Speech language and communication*. New York: Academic Press; 1995. p. 217-262.
- Tomasello M. The social bases of language acquisition. *Social Development*. 1992; 1(1):67–87.
- Tomasello, M. *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press; 2003.
- Tomasello M, Barton M. Learning words in nonostensive contexts. *Developmental Psychology*. 1994; 30:639–650.
- Tomasello M, Ferrar MJ. Joint attention and early language. *Child Development*. 1986; 57:1454–1463. [PubMed: 3802971]
- Tomasello M, Kruger AC. Joint attention on actions: Acquiring verbs in ostensive and non-ostensive contexts. *Journal of Child Language*. 1992; 19(2):311–333. [PubMed: 1527205]
- Tomasello, M.; Merriman, WE., editors. *Beyond words for things: Acquisition of the verb lexicon*. New York: Academic Press; 1995.
- Trueswell, JC. Using eye movements as a developmental measure within psycholinguistics. In: Sekerina, IA.; Fernández, EM.; Clahsen, H., editors. *Language processing in children*. Amsterdam: John Benjamins; 2008. p. 73-96.
- Trueswell, J.; Gleitman, LR. Children's eye movements during listening: Evidence for a constraint-based theory of parsing and word learning. In: Henderson, JM.; Ferreira, F., editors. *Interface of language, vision, and action: Eye movements and the visual world*. New York: Psychology Press; 2004. p. 319-346.
- Trueswell, JC.; Gleitman, LR. Learning to parse and its implications for language acquisition. In: Gaskell, G., editor. *Oxford handbook of psycholinguistics*. Oxford, UK: Oxford University Press; 2007. p. 635-656.
- Trueswell JC, Sekerina I, Hill NM, Logrip ML. The kindergarten-path effect: Studying on-line sentence processing in young children. *Cognition*. 1999; 73:89–134. [PubMed: 10580160]

- Trueswell, J.; Tanenhaus, M. Toward a lexicalist framework of constraint-based syntactic ambiguity resolution. In: Clifton, C.; Rayner, K.; Frazier, L., editors. *Perspectives on sentence processing*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.; 1994. p. 155-179.
- Tversky A. Features of similarity. *Psychological Review*. 1977; 84(4):327–350.
- Von Berger E, Wulfeck B, Bates E, Fink N. Developmental changes in real-time sentence processing. *First Language*. 1996; 16(47):193–222.
- Wittgenstein, L. *Philosophical investigations*. Oxford, UK: Basil Blackwell; 1953. (G. E. M. Anscombe, Trans.).
- Woodward AL. Infants selectively encode the goal object of an actor’s reach. *Cognition*. 1998; 69:1–34. [PubMed: 9871370]
- Woodward AL, Sommerville JA. Twelve-month-old infants interpret action in context. *Psychological Science*. 2000; 11:73–77. [PubMed: 11228848]
- Xu F, Tenenbaum JB. Sensitivity to sampling in Bayesian word learning. *Developmental Science*. 2007a; 10:288–297. [PubMed: 17444970]
- Xu F, Tenenbaum JB. Word learning as Bayesian inference. *Psychological Review*. 2007b; 114(2): 245–272. [PubMed: 17500627]
- Yantis S, Jonides J. Abrupt visual onsets and selective attention: Voluntary versus automatic allocation. *Journal of Experimental Psychology: Human Perception and Performance*. 1990; 16:121–134. [PubMed: 2137514]
- Yantis S, Jonides J. Attentional capture by abrupt onsets: New perceptual objects or visual masking? *Journal of Experimental Psychology: Human Perception and Performance*. 1996; 22:1505–1513. [PubMed: 8953232]
- Yu C, Smith LB. Rapid word learning under uncertainty via cross-situational statistics. *Psychological Science*. 2007; 18(5):414–420. [PubMed: 17576281]
- Yuan S, Fisher C. “Really? She blicked the baby?”: Two-year-olds learn combinatorial facts about verbs by listening. *Psychological Science*. 2009; 20(5):619–626. [PubMed: 19476591]



a.



b.

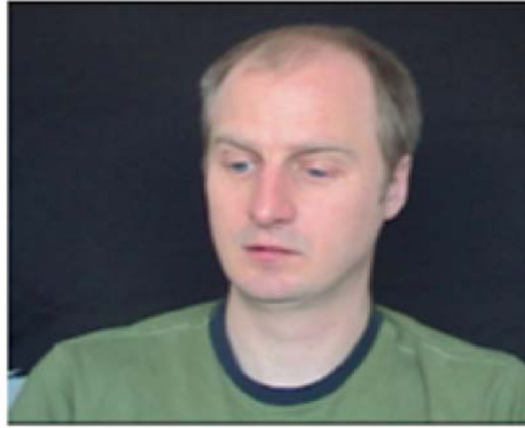
FIGURE 1.

Example of *chase/flee* contexts. In 1a., the rabbit is chasing the elephant, but the elephant also is fleeing the rabbit. In 1b., the elephant is chasing the rabbit, but the rabbit also is fleeing the elephant.



FIGURE 2.

Warm-up trial as seen from the child's perspective in the experiment. The actor looked directly down at the characters and said either, "They're mooping" (Exp. 1) or "The man and woman are mooping" (Exp. 2).

**FIGURE 3.**

Example screenshot from a target video. In this example, Andy is looking at Character B. Andy then says “*He’s blicking him*” (Exp. 1) or “*The rabbit is blicking the elephant*” or “*The elephant is blicking the rabbit*” (Exp. 2).

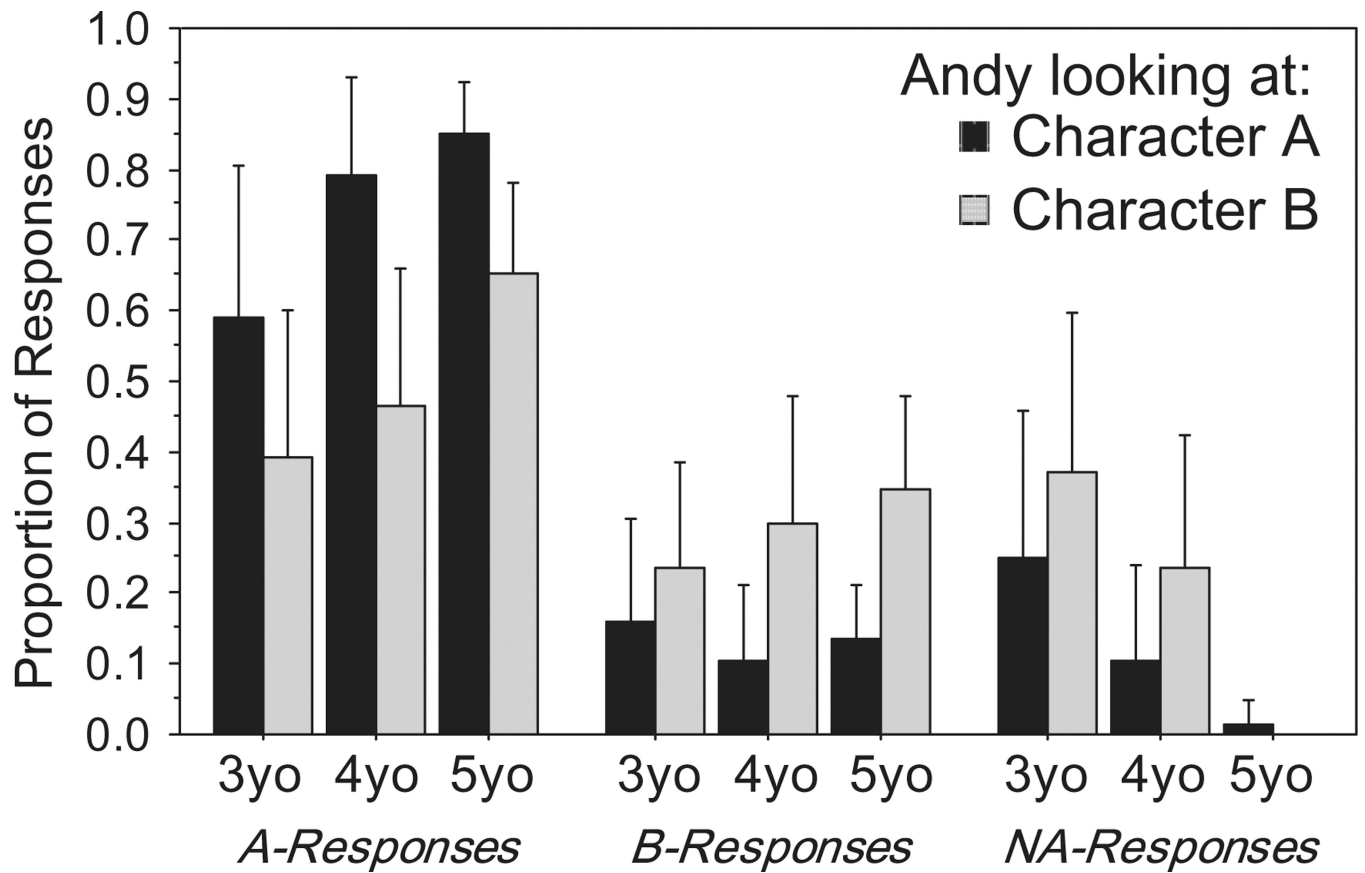


FIGURE 4. Mean proportion of A-, B- and NA-responses (Experiment 1). Error bars indicate ± 1 Standard Error (Exp. 1).

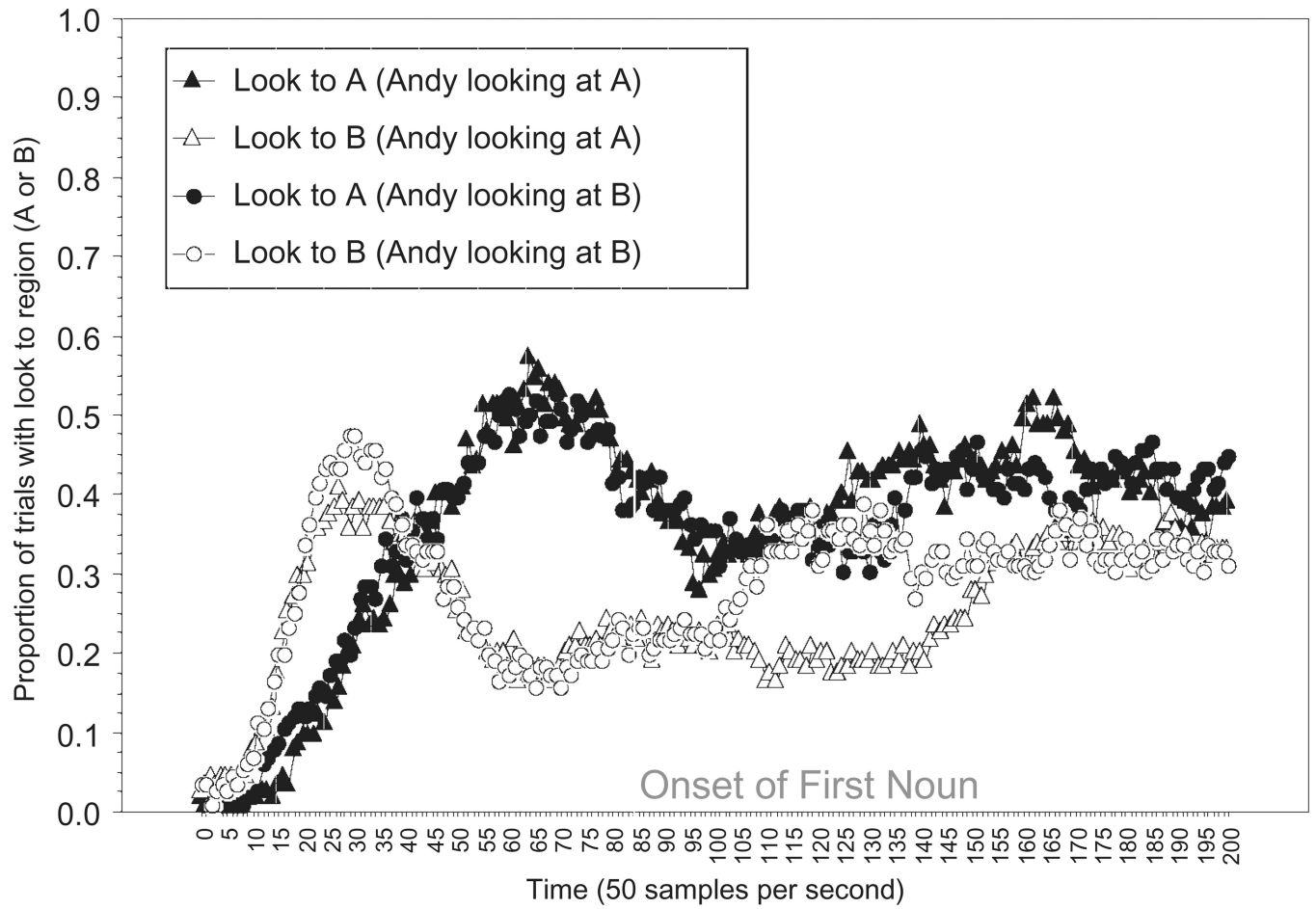


FIGURE 5. Time course of looks to Character A and to Character B as a function of Andy’s direction of Gaze (Exp. 1).

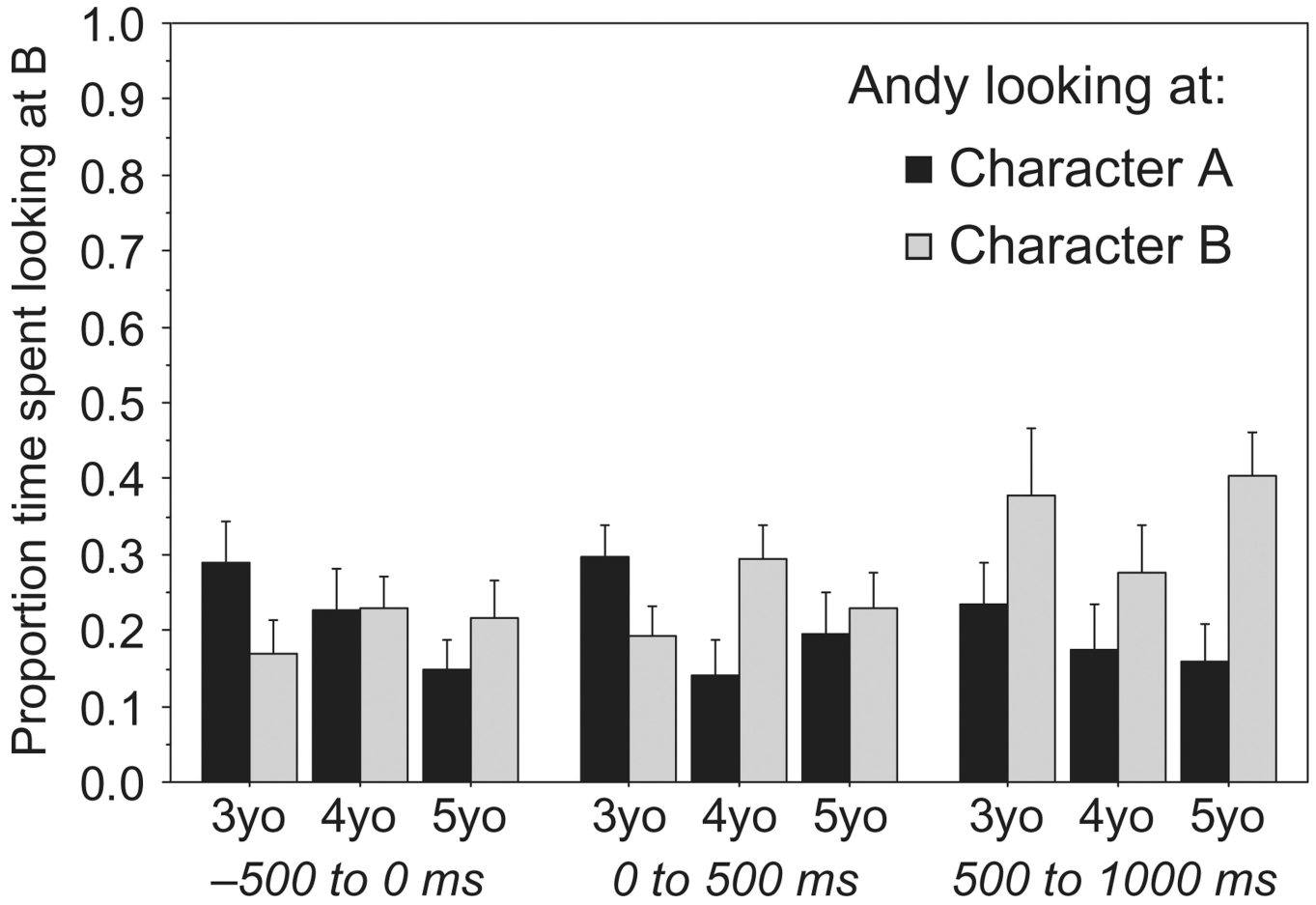


FIGURE 6. Proportion of time spent looking at Character B as a function of Andy’s direction of Gaze, for three 500 ms time windows around utterance onset (Exp. 1).

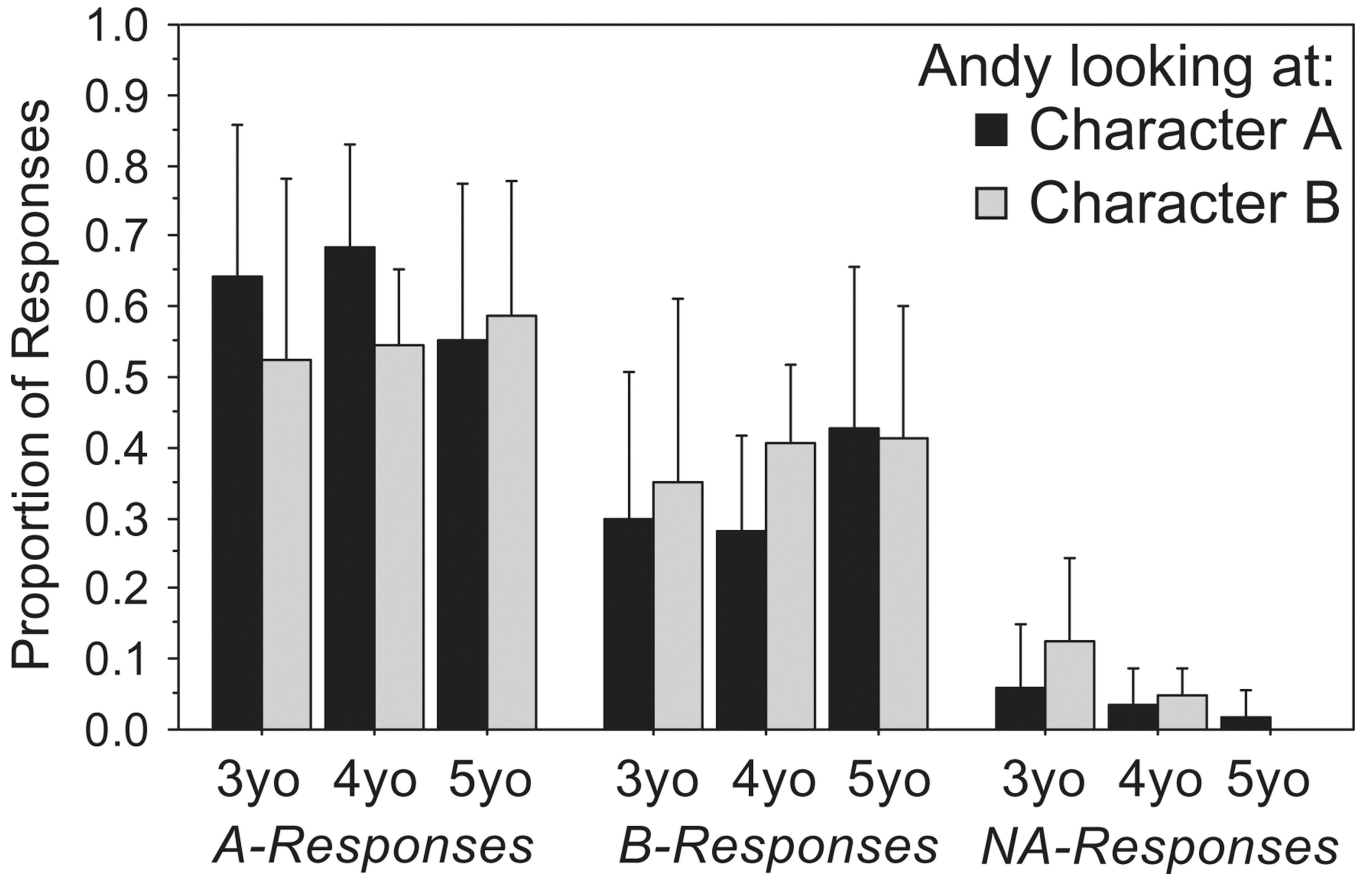


FIGURE 7. Mean proportion of A-, B- and NA-responses as a function of Andy’s direction of Gaze, collapsing across Syntax conditions. Error bars indicate ± 1 Standard Error (Exp. 2).

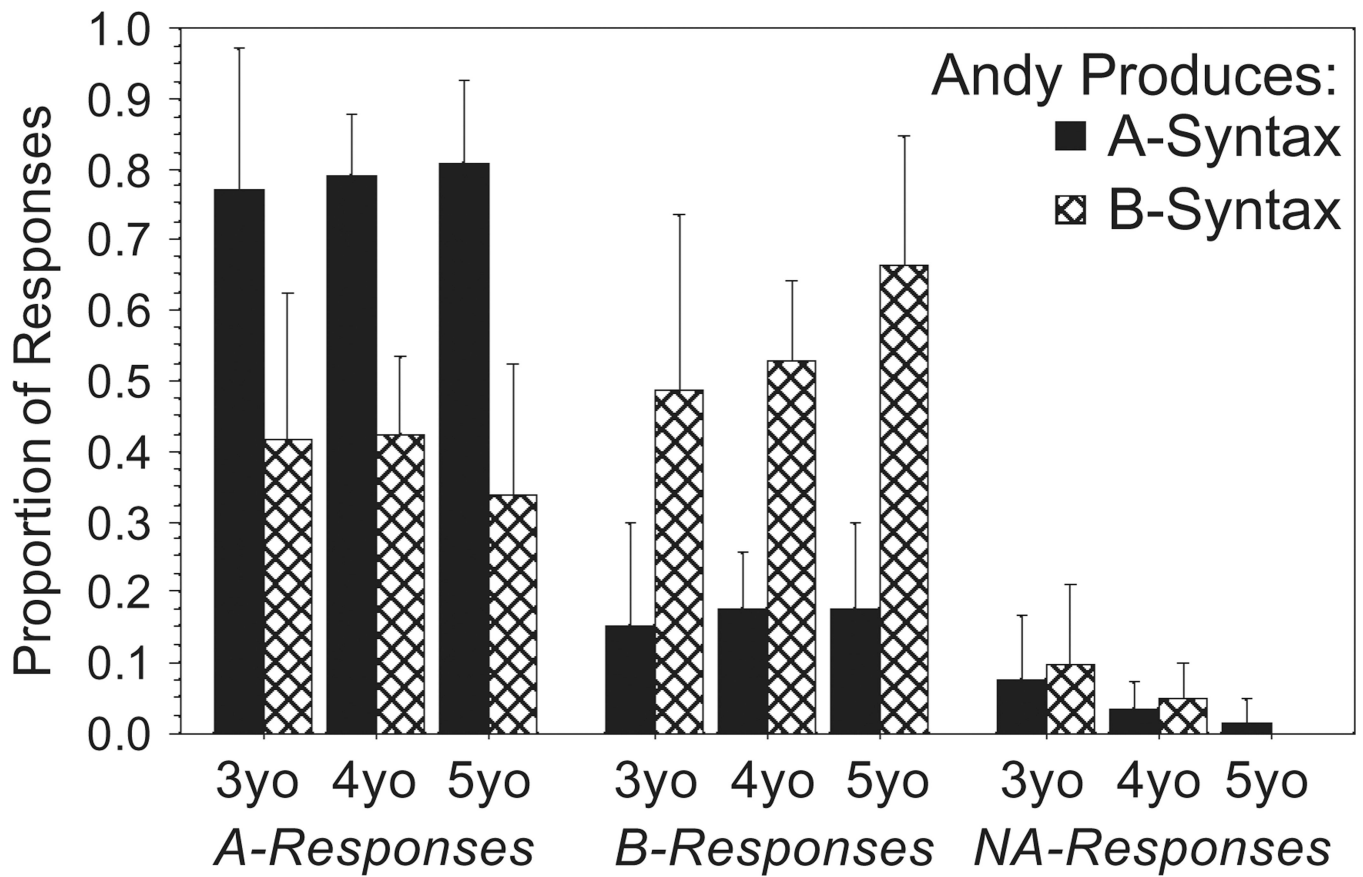


FIGURE 8. Mean proportion of A-, B- and NA-responses as a function of Andy's Syntax, collapsing across Gaze conditions. Error bars indicate ± 1 Standard Error (Exp. 2).

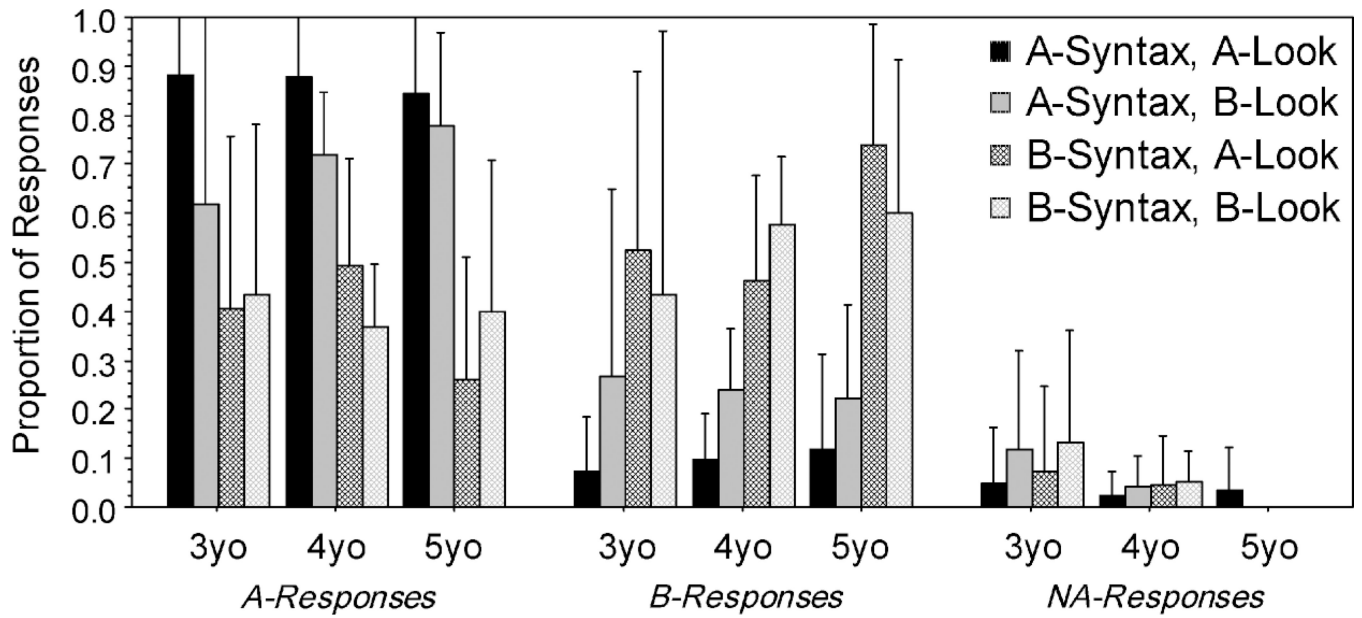


FIGURE 9. Mean proportion of A-, B- and NA-responses as a function of Andy's direction of Gaze and Andy's Syntax. Error bars indicate ± 1 Standard Error (Exp. 2).

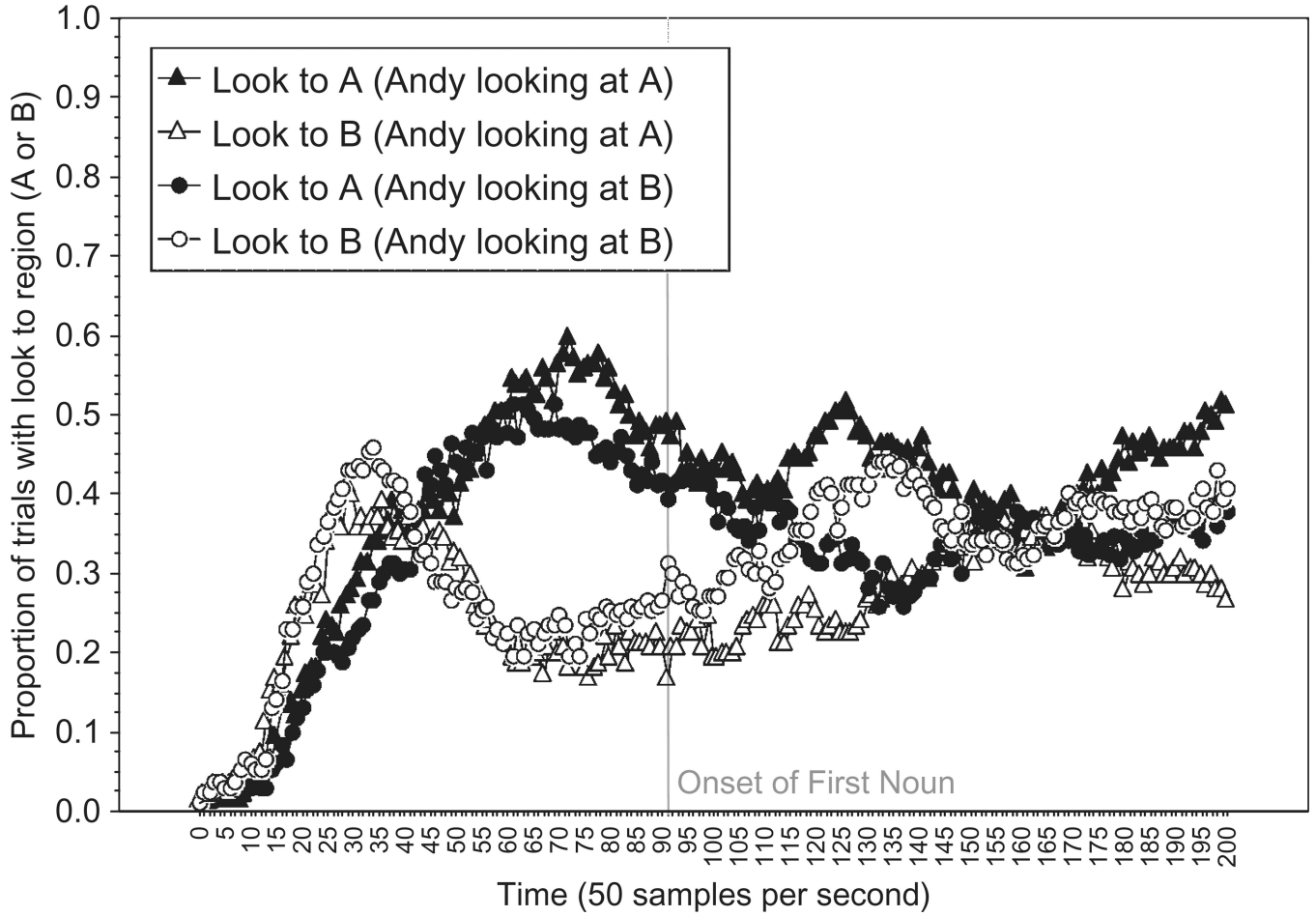


FIGURE 10. Time course of looks to Character A and to Character B as a function of Andy’s direction of Gaze, collapsing across Syntax conditions (Exp. 2).

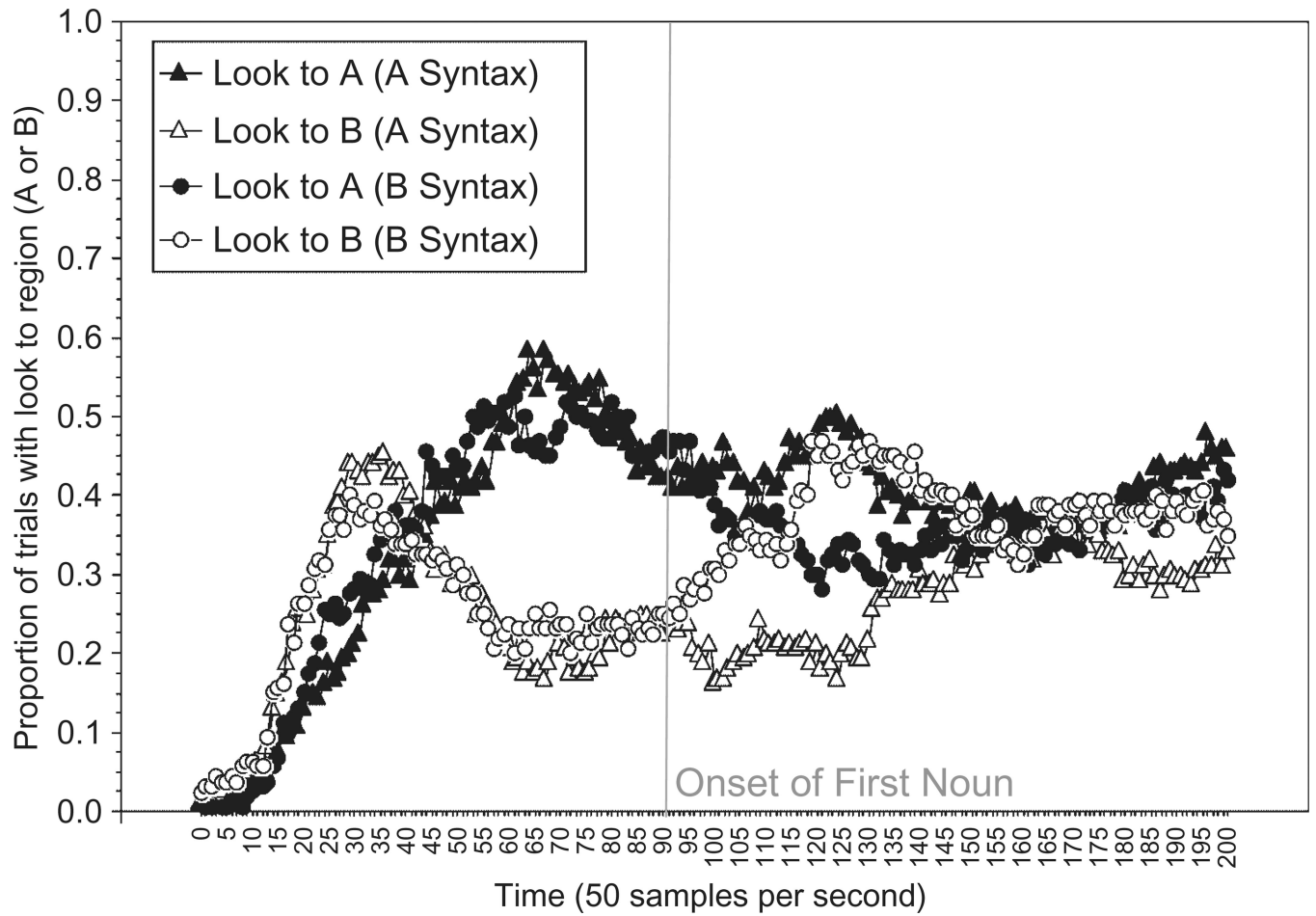
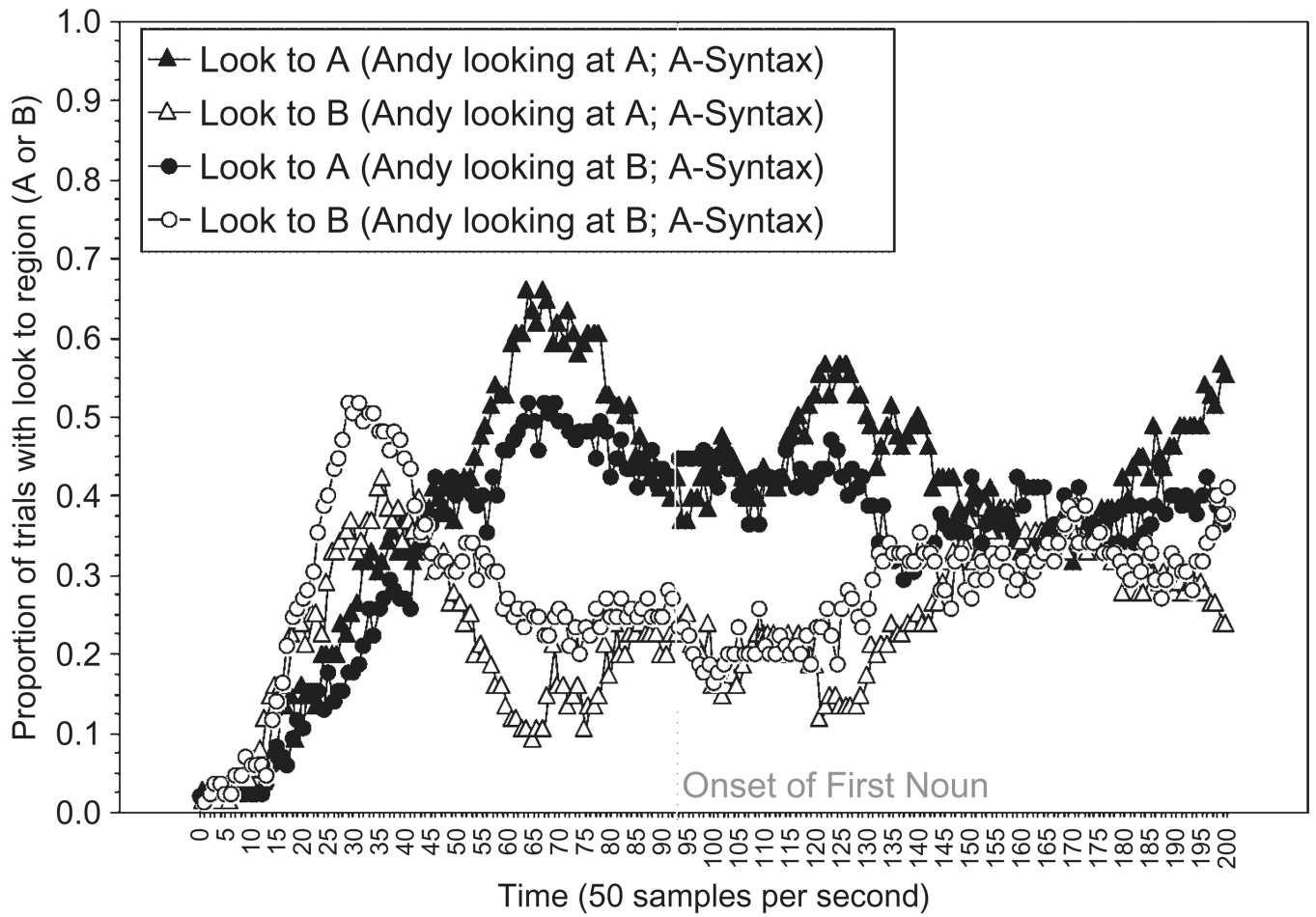
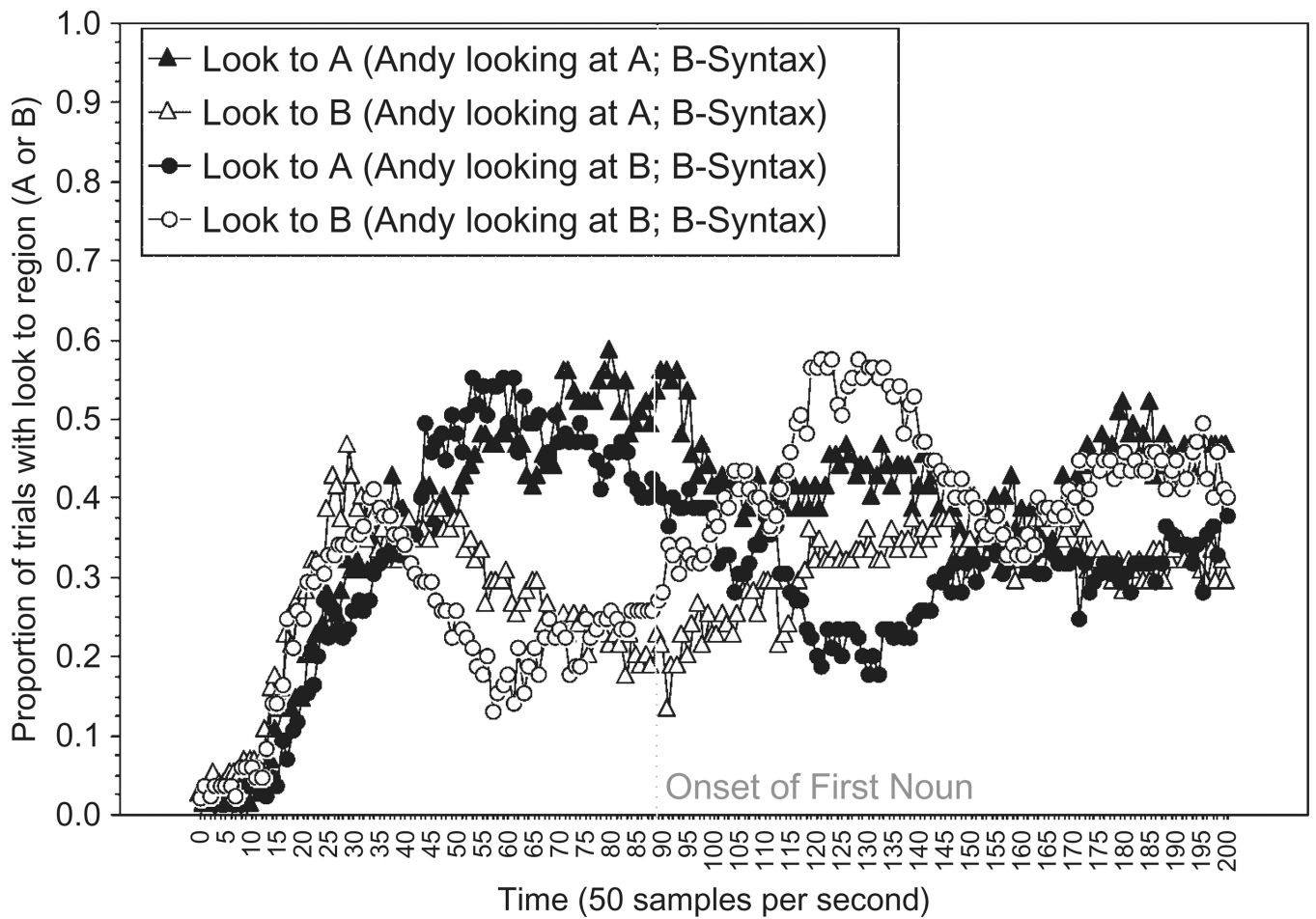


FIGURE 11. Time course of looks to Character A and to Character B as a function of Andy’s Syntax, collapsing across Gaze conditions (Exp. 2).



A



B

FIGURE 12.

A Time course of looks to Character A and to Character B as a function of Andy’s direction of Gaze. A-Syntax condition only (Exp. 2).

B Time course of looks to Character A and to Character B as a function of Andy’s direction of Gaze. B-Syntax condition only (Exp. 2).

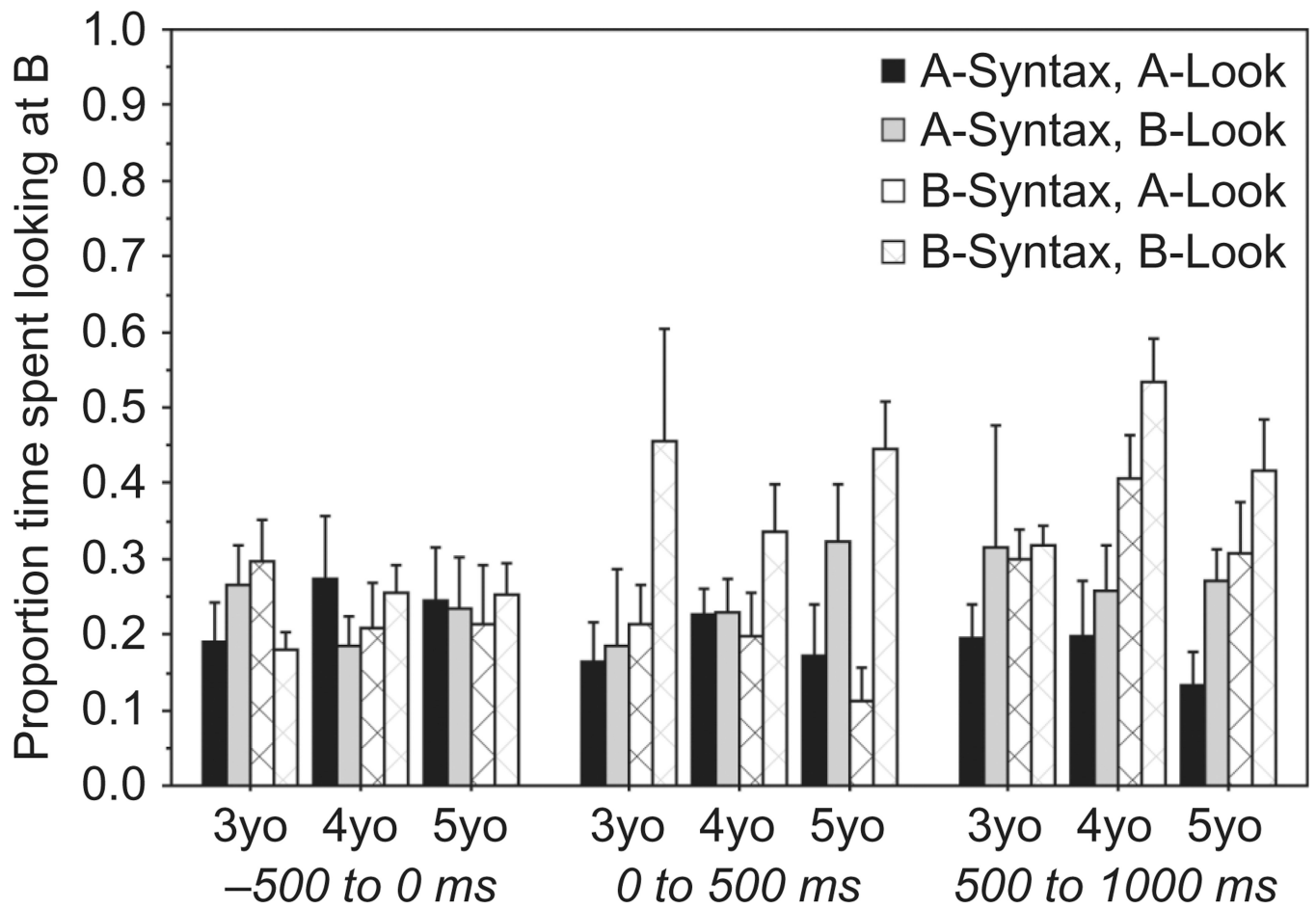


FIGURE 13. Proportion of time spent looking at Character B as a function of Andy's direction of Gaze and Andy's Syntax, for three 500 ms time windows around utterance onset (Exp. 2).

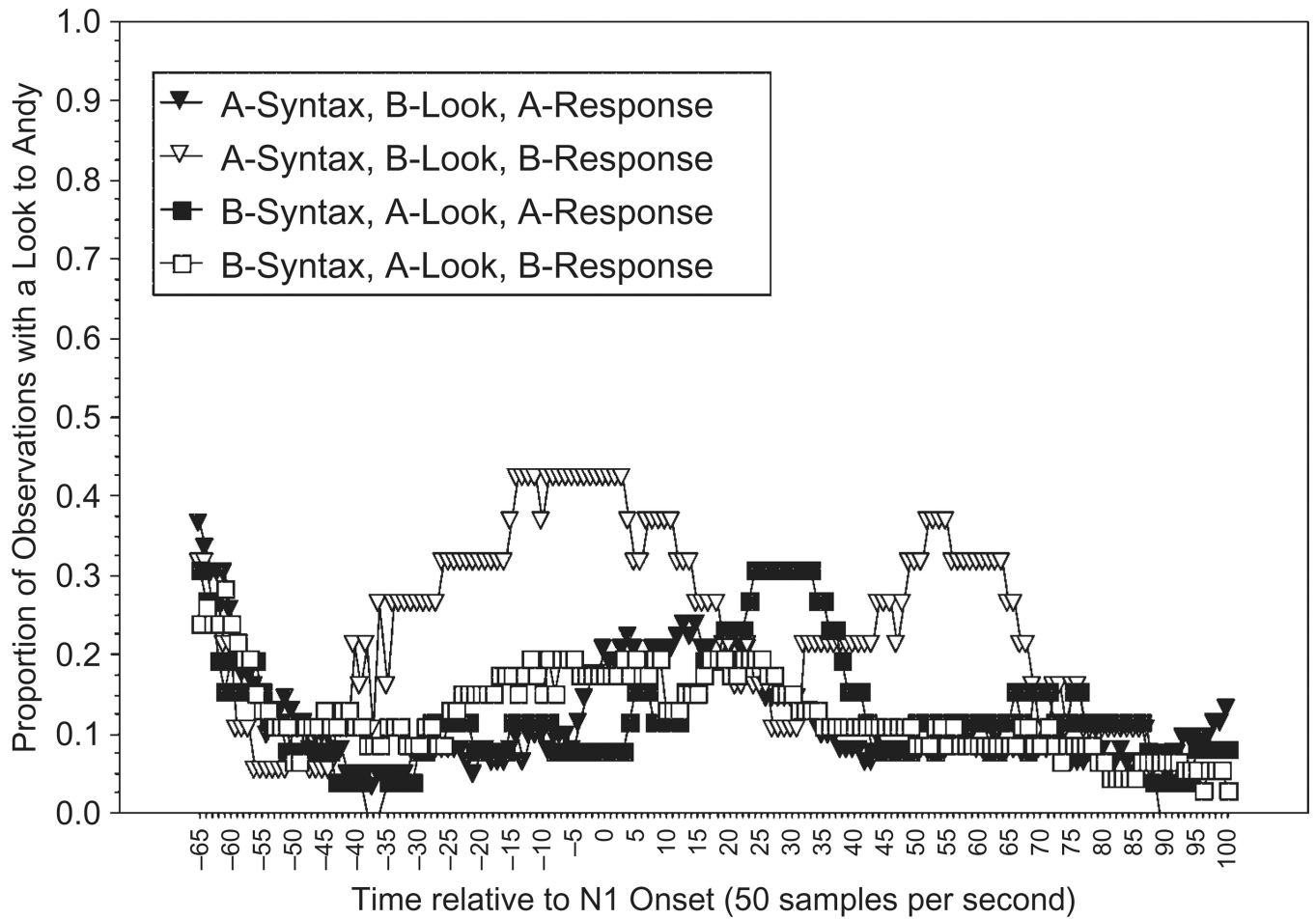


FIGURE 14. Time course of participants' looks to the speaker, Andy, as a function of Andy's Syntax and Andy's direction of Gaze (cue-conflict conditions only) (Exp. 2).

TABLE 1
ANOVA Results for the Mean Proportion of A-responses, Reported Separately for Subject (F_1) and Item (F_2) Means

Effect	F_1	df_1	p_1	F_2	df_2	p_2
Experiment 1						
Andy's Gaze	16.20	1, 36	<0.0005	18.71	1, 6	<0.01
Age	6.83	2, 36	<0.005	7.22	2, 12	<0.01
Gaze*Age	0.51	2, 36	-	2.22	2, 12	-
Experiment 2						
Andy's Gaze	2.38	1, 47	-	2.76	1, 6	-
Andy's Syntax	37.56	1, 47	<0.0001	14.88	1, 6	<0.005
Age	0.37	2, 47	-	1.10	2, 12	-
Gaze*Syntax	1.91	1, 47	-	9.50	1, 6	<0.05
Gaze*Age	1.41	2, 47	-	1.65	2, 12	-
Syntax*Age	0.47	2, 47	-	1.19	2, 12	-
Gaze*Syntax*Age	0.39	2, 47	-	0.47	2, 12	-

TABLE 2

ANOVA Results for the Mean Proportion of B-looks Relative to Andy's Utterance Onset, Reported Separately for Subject (F_1) and Item (F_2) Means, Experiment 1

Effect	F_1	df_1	p_1	F_2	df_2	p_2
<i>-500 to 0 ms</i>						
Andy's Gaze	0.16	1, 33	-	0.26	1, 6	-
Age	0.71	2, 33	-	0.46	2, 12	-
Gaze*Age	1.69	2, 33	-	1.17	2, 12	-
<i>0 to 500 ms</i>						
Andy's Gaze	0.60	1, 33	-	0.49	1, 6	-
Age	0.17	2, 33	-	0.20	2, 12	-
Gaze*Age	4.24	2, 33	<0.05	5.61	2, 12	<0.05
<i>500 to 1000 ms</i>						
Andy's Gaze	12.37	1, 33	<0.005	8.21	1, 6	<0.05
Age	0.73	2, 33	-	0.79	2, 12	-
Gaze*Age	0.98	1, 33	-	0.32	2, 12	-

TABLE 3

ANOVA Results for the Mean Proportion of B-looks Relative to Andy's Utterance Onset, Reported Separately for Subject (F_1) and Item (F_2) Means, Experiment 2

Effect	F_1	df_1	p_1	F_2	df_2	p_2
<i>-500 to 0 ms</i>						
Andy's Gaze	0.05	1, 43	-	0.12	1, 6	-
Andy's Syntax	0.01	1, 43	-	0.07	1, 6	-
Age	0.01	2, 43	-	0.55	2, 12	-
Gaze*Syntax	0.01	1, 43	-	0.03	1, 6	-
Gaze*Age	0.11	2, 43	-	0.02	2, 12	-
Syntax*Age	0.02	2, 43	-	0.05	2, 12	-
Gaze*Syntax*Age	1.70	2, 43	-	3.55	2, 12	0.06
<i>0 to 500 ms</i>						
Andy's Gaze	13.75	1, 43	<0.001	27.99	1, 6	<0.005
Andy's Syntax	4.01	1, 43	0.05	2.63	1, 6	-
Age	0.07	2, 43	-	0.62	2, 12	-
Gaze*Syntax	5.57	1, 43	<0.05	4.40	1, 6	0.08
Gaze*Age	2.04	2, 43	-	2.89	2, 12	-
Syntax*Age	0.95	2, 43	-	1.26	2, 12	-
Gaze*Syntax*Age	0.11	2, 43	-	0.20	2, 12	-
<i>500 to 1000 ms</i>						
Andy's Gaze	7.22	1, 43	<0.05	3.64	1, 6	0.11
Andy's Syntax	11.50	1, 43	<0.005	23.72	1, 6	<0.005
Age	2.00	2, 43	-	1.85	2, 12	-
Gaze*Syntax	0.62	1, 43	-	0.01	1, 6	-
Gaze*Age	0.18	2, 43	-	0.04	2, 12	-
Syntax*Age	0.06	2, 43	-	0.52	2, 12	-
Gaze*Syntax*Age	0.31	2, 43	-	0.32	2, 12	-