RUNNING HEAD: Effects of Psychological Attention

Effects of Psychological Attention on Pronoun Comprehension

Jennifer E. Arnold and Shin-Yi C. Lao

Department of Psychology, University of North Carolina at Chapel Hill

Address all correspondence to: Jennifer Arnold, UNC Chapel Hill, Dept. of Psychology, Davie Hall #337B, CB #3270, Chapel Hill, NC 27599-3270, Email: [jarnold@email.unc.edu](mailto:jarnold@email.unc.edu); 919-843-5737

**Abstract**

Pronoun comprehension is facilitated for referents that are focused in the discourse context. Discourse focus has been described as a function of attention, especially shared attention, but few studies have explicitly tested this idea. Two experiments used an exogenous capture cue paradigm to demonstrate that listeners’ visual attention at the onset of a story influences their preferences during pronoun resolution later in the story. In both experiments trial-initial attention modulated listener’s transitory biases while considering referents for the pronoun, whether it was in response to the capture cue or not. These biases even had a small influence on listeners’ final interpretation of the pronoun. These results provide independently-motivated evidence that attentional mechanisms impact the on-line processes of pronoun comprehension. Trial-initial attentional shifts were made on the basis of non-shared, private information, demonstrating that attentional effects on pronoun comprehension are not restricted to shared attention among interlocutors.

Keywords: pronoun resolution, attention, discourse processing, accessibility

Understanding language frequently requires listeners to find referents for pronouns, for example in sentences like *Paige went skiing with Athena, and she fell down.* Listeners typically do so quite rapidly, despite the frequent ambiguity of pronouns, in part because they can draw on extra-linguistic biases based on the previous discourse and nonlinguistic context. Some entities (e.g., recently or prominently mentioned ones) are perceived as better referents. These are called **focused** or **salient** in the discourse, and are typically more accessible during the interpretation of subsequent referring expressions (Ariel, 1990; Bock & Irwin, 1980; Chafe, 1994; Grosz, Joshi, & Weinstein, 1995; Gundel, Hedberg, & Zacharaski, 1993; see Arnold, 1998, 2008, 2010 for reviews). The link between cognitive status and pronoun comprehension is widely accepted, but there are many open question about the mechanisms by which information becomes mentally privileged, and how this status affects pronoun comprehension.

One view suggests that is that some information is represented in memory in such a way that it is easier to access it, and that this facilitates the comprehension of reference, in particular underspecified forms like pronouns. This mental status has been described as “salience”, “accessibility”, “activation”, “givenness”, “topicality” or “prominence” (for example, Ariel, 1990. 2001; Arnold, 1998, 2008, 2010; Bower & Morrow, 1990; Brennan, 1995; Chafe, 1994; Garnham, Traxler, Oakhill, & Gernsbacher, 1996; Gundel, Hedberg, & Zacharski, 1993; Givon, 1983; Grosz, Joshi, & Weinstein, 1995; Kaiser & Trueswell, 2004). Although existing proposals differ in important details, they share the use of nonlinguistic mental representations as explanations for both speakers’ choices in production, and listeners’ preferences in comprehension. This mental status is often assumed to be a gradient representation (e.g., Ariel, 1990; Arnold, 1998; Chafe, 1994), although many theories suggest that there is also a single most highly focused referent (Gundel, Hedberg, & Zacharski, 1993; Givon, 1983; Grosz, Joshi, & Weinstein, 1995; Stevenson et al., 2000). For example, following the sentence *Sophie jumped high on the trampline*, the character *Sophie* is highly salient, and the most likely referent of a matching pronoun, e.g. *…and she laughed.* By contrast, entities that are unrelated to the current situation are extremely low in salience, and are unlikely to be considered as potential referents for pronouns.

The question we ask here is whether pronoun comprehension is influenced by the listener’s attention. This is a question worth asking, because there is uncertainty in the literature about whether the mechanism behind pronoun comprehension relies on actual, psychological attention, or by a language-specific category that is called “in focus”. On one hand, some scholars discuss referential salience in terms of the reference comprehension as it relates to the speaker’s assumptions about the listener’s attention (Gundel et al., 1993; Chafe, 1994) or memory (Ariel, 1990, 2001). On the other hand, much of the literature uses a more discourse-specific notion, termed simply “the focus” (e.g., Dahan et al., 2002; McKoon, Greene, & Ratcliff, 1993; Marslen-Wilson, Tyler, & Koster, 1993; Stephenson, Crawley, & Kleinman, 1994). The implication is that “in focus” information is also that which is attended by the discourse participants. Although researchers are specific about this link, in some cases the link between attention and discourse focus is not explicit. Moreover, there have been no explicit tests of whether attention is actually involved in the representation of discourse entities, or how. Is it that pronoun comprehension is driven by a language-specific category that we call “focus”, or by actual fluctuations in attention?

The role of attention in reference comprehension is complex, because it is intertwined with known effects of the discourse context. It is well established that reference comprehension is highly constrained by the discourse context. Listeners tend to perceive as accessible those things that were recently mentioned, especially those mentioned in prominent linguistic positions like subject or first-mentioned position (Ariel, 1990; Arnold, Eisenband, Brown-Schmidt, & Trueswell, 2000; Chafe, 1976; Clark & Sengul, 1979; Givón, 1983, Brennan, 1995; Gordon, Grosz, & Gilliom, 1993), or the focus of a cleft (Almor, 1999; Arnold, 1998; Cowles, Walenski, & Kleunder, 2007; c.f. Foraker & McElree, 2007). The effects of discourse status are especially apparent in the comprehension of pronouns, which are lexically underspecified (Arnold et al., 2000).

Yet the causal relationship between discourse focus and attention is unclear. The discourse context alone is likely to influence listeners’ attention. Prominent entities in a discourse are perceived as topical (Givon, 1983) and expected (Arnold, 1998; Korneef & van Berkum, 2006; Tily & Piantadosi), and thus likely to draw attention. That is, the kinds of discourse contexts that facilitate pronoun comprehension are also the kinds of contexts that direct attention to the pronoun referent. Which, then, guides pronoun comprehension – the context or the attended state of some referents? Despite the frequent use of attention to explain discourse status effects on during processing, there has been little direct assessment of its psychological role.

In this paper, we address this question in two experiments, which examine the influence of attention on pronoun comprehension by adopting an attentional capture manipulation from research on visual attention (Folk, Remington, & Johnson, 1992; Jonides, 1981; Müller & Rabbit, 1989). In particular, we examine the role of individual, non-shared attention in the context of a public discourse. We focus on listeners’ immediate interpretations of the pronoun, as it is encountered and immediately afterward – that is, on-line processing. We expect that in this domain we are most likely to observe any effects of attention over and above discourse biases. These on-line biases are then compared to measures of final, off-line interpretations of the pronoun.

**How attention relates to referential accessibility**

The use of a non-discourse manipulation of attention is important, because it is unclear how attentional mechanisms may be related to the effects of the discourse itself on referent accessibility. We begin by considering some potential relationships. Here we use the term “referential accessibility” to represent the privileged cognitive status (aka salience, prominence, etc.) that is predicted to facilitate pronoun comprehension.

***Accessibility as shared attention.***On one hand, the linguistic context is a public record of the current task, which is consistent with the broader assumption that referential accessibility is a function of shared attention specifically, and not just the egocentric attention of each discourse participant (e.g., Clark, 1996). Here we use the term **shared attention** to denote information that one person considers likely to be attended by their discourse partner – that is, the attended portion of common ground. This is a somewhat weaker definition than **joint attention**, which may require both participants to be aware that the other is attending (Tomasello, 1995). Indeed, many theorists emphasize the importance of shared attention for describing referent accessibility. For example, Gundel et al.’s (1993) Givenness Hierarchy explains the cognitive status of discourse entities in terms of “assumptions that a cooperative speaker can reasonably make regarding the addressee’s knowledge and attention state” (p. 275; see also Bard, Anderson & Sotillo, 2000; Brennan, 1995; Chafe, 1994; Levelt, 1989). This approach suggests that reference processing is only affected by evidence about the attention of one’s interlocutor (see also Clark & Marshall, 1981). For example, if you are daydreaming about a chocolate bar at the moment that a student hands you a paper and says “Is it good?”, you should not interpret “it” as the chocolate bar. The discourse record is a good estimate of what is mutually attended by discourse participants. But non-linguistic information, like eye gaze, can also signal shared attention (Hanna & Brennan, 2008; Nappa & Arnold, under review; Goodrich & Hudson Kam, 2008, 2009). You could imagine watching a juggler perform on the street and saying “Wow, he’s amazing” to the stranger next to you. The pronoun would be fully interpretable, based on evidence of physical co-presence, and physical cues like posture and eyegaze that indicate shared attention on the juggler -- even if the juggler had not been previously mentioned.

***Accessibility as egocentric attention.*** On the other hand, some researchers emphasize the attention of the comprehender. Bower and Morrow (1990; see also Morrow, Bower, & Greenspan, 1987, 1989) suggest that readers use mental models to represent the objects and characters in a discourse, and as they move through the discourse their attention shifts to the physical location in their mental representation. This raises the possibility that it is the individual, non-joint attention that guides accessibility. Likewise, discourse cues like first-mentioned status or recency could directly modulate the listeners’ focus of attention without requiring representations of joint attention (Arnold, 2008).

***Accessibility as a language-specific construct.*** A final possibility is that discourse information affects reference processing because there are discourse-specific rules that make some entities preferred as referents (e.g., first-mentioned ones). In fact, there must be some contribution of language-specific rules, since language like Finnish have different pronouns that respond to different aspects of the linguistic context (Kaiser & Trueswell, 2008). More generally, the textual information in the discourse goes a long way toward predicting the preferred referent for a pronoun (Arnold, 1998). Centering Theory is a computational model that aims to use the text to make such predictions (Grosz, et al., 1995; also see Walker, Prince, & Joshi 1998 and papers within). In Centering, the highest-ranked entity referred to in the next utterance is the “center of attention.” But it is unclear whether “center” is just a discourse category, or a description of a psychological state of the discourse participants. In fact, Grosz and Sidner (1986) state explicitly that in their model (a precursor to Centering theory), “The attentional state is a property of the discourse itself, not of the discourse participants.” (p. 179). While this model implies that there is a relationship between discourse focus and psychological attention, the proposed mechanism for representing discourse status is an abstraction.

Together, the variability in how accessibility is discussed in the literature raises questions about the role of attention. Does psychological attention guide the assessment and representation of discourse accessibility? There are many ways in which it could do so: attention could be only one determinant of a separate representation of discourse accessibility, or it could be that attention entirely determines referential accessibility. An alternate possibility is that pronoun comprehension is guided only by the discourse context, for example the well-known subject bias. On this account, attention in and of itself does not influence accessibility, even though the linguistic context may also influence attention.

Thus, our goal is to investigate whether attentional shifts can influence referential accessibility. However, to do so we cannot simply examine the effects of the linguistic context, because as we have argued, these could guide reference comprehension either directly, or by affecting attention to discourse elements. We therefore turn to a domain in which attention is relatively well understood: visual attention.

**Does visual attention impact referential accessibility?**

Visual attention offers a promising window onto questions about the role of attention reference processing. We are not suggesting that visual attention is isomorphic with discourse focus. Indeed, we will describe several ways in which visual attention has different properties than discourse focus / accessibility. Nevertheless, there good reason to suspect that visual attention overlaps with attentional processes during language use.

One potential concern is that linguistic attention and visual attention have different effects. The “discourse focus” (or “center of attention”) is typically used to describe backgrounded, given, topical referents -- i.e., those that are highly expected to be mentioned again. In contrast, visual attention is often attracted to objects that are new and need further processing, (see Bock, Irwin, & Davidson, 2004, for discussion) – e.g., those that appear suddenly (e.g. Jonides & Yantis, 1988; Remington, Johnston & Yantis, 1992; Yantis, 1993), are unfamiliar (e.g. Brockmole & Boot, 2009; Theeuwes, 1994), or do not fit thematically with the scene (De Graef, Christiaens, & d’Ydewalle, 1990; Henderson, Weeks, & Hollingworth., 1999; Lofuts & Mackworth, 1978). For example, van der Muelen, Meyer, & Levelt (2001) found that speakers looked at given referents in a display less than new referents, and they spent less time looking at the object when preparing a pronoun than a name. That is, they found that visual attention was enhanced for the objects with low referential accessibility.

Nevertheless, visual attention is not a unitary mechanism. Attention can be driven by both **endogenous** orienting, i.e. following task-relevant goals, and **exogenous,** stimulus-driven orienting (Folk, et al., 1992; Jonides, 1981; Müller & Rabbit, 1989, Theeuwes, 1991)[[1]](#footnote-1). Although endogenous and exogenous attention have distinct properties (Berger, Henik, & Rafal, 2005; Hopfinger & Ries, 2005; Jonides, 1981), there is both neural and functional overlap (Gleitman, January, Nappa, & Trueswell, 2007; Hopfinger & Mangun, 1998). While exogenous, involuntary attention is typically associated with sudden onset, i.e. “new” items, Chanon and Hopfinger (2008) found that memory for a scene directed attention towards previously-viewed (“old”) items; this effect was not modulated by task instructions, suggesting that it was unintentional. The stimulus-driven effects of attentional capture are additionally modulated by top-down attentional control settings, e.g. where color contrast attracts attention only if relevant for the current task (Folk et al., 1992; see also Yantis & Jonides, 1990). Thus, visual attention can be automatically directed towards both familiar and novel stimuli, and is controlled by both goal-directed (endogenous) and stimulus-directed (exogenous) mechanisms, and the interaction between them (but see Berger, et al., 2005). Given the overlap between these distinct types of visual attention, it is possible that similar overlap exists between exogenous visual orienting and attention to linguistically mentioned discourse entities.

There is also some limited evidence for overlap between visual attention and linguistic processing in the literature. We know that linguistic processing influences visual attention (Salverda & Altmann, 2005; see also Altmann & Kamide, 2007). Salverda and Altmann (2005) found that hearing a reference to a pictured object slowed the execution of a saccade toward an exogenous visual capture cue. Here we ask whether the reverse also hold true. That is, can visual attention affect accessibility during linguistic referential processing?

There is reason to believe that visual attention might impact speakers’ choices during reference production, given research on the speaker’s choices about how to begin an utterance. For example, speakers have choices about whether to begin a sentence with the thematic Actor or Undergoer, .e.g, the man or the dog in *The dog chased the man* or *The man was chased by the dog* (Bock et al., 2004; Gleitman et al., 2007; Tomlin, 1997). Speakers tend to choose syntactic structures that allow them to begin their utterances with information that is more accessible conceptually or lexically (Bock, 1982; 1986; Ferreira & Dell, 2000; Ferreira, 2003). This accessibility is influenced by the same kinds of things as referential accessibility, e.g. the preference to order given before new (Arnold, Wasow, Losongco, & Ginstrom, 2000; Hawkins, 1994; Bock & Irwin, 1980).

Evidence for the link between visual attention and language also comes from findings that speakers tend to choose structures that allow them to begin their utterance with the referent that they initially fixated in a scene (Gleitman et al., 2007). Even more striking, Gleitman et al. were able to modulate the speaker's choice of starting point by manipulating their attention with an exogenous attentional capture cue. Speakers described scenes with two characters. Immediately before the picture appeared, a black square appeared for 80 ms, which increased the likelihood of starting the sentence with the cued character. These results suggest that visual and linguistic attention overlap at some level, despite their differences. On the other hand, conflicting evidence comes from Bock et al. (2004), who monitored participants' eye movements as they told time from an analog clock. Speakers looked at the minutes hand more often when using a minutes-first time frame (e.g., *ten to two*), and at the hour hand more when using an hour-first time frame (*one-fifty*). However, their choice of starting point was not reflected in their initial fixation to the scene, suggesting that visual attention was driven by planning processes, and not the other way around. Together these studies suggest that the links between visual attention and referential accessibility may go either direction, or both.

Nevertheless, evidence that egocentric focus of attention influences production does not necessarily mean that it also affects accessibility during comprehension. The speaker’s goal during language production is to communicate their own point of view in such a way that the listener understands. For this task the speaker’s personal focus of attention is likely to be a part of what they are trying to communicate. But comprehension is aimed at uncovering the speaker’s intentions, and the speaker’s point of view. Thus, attentional processes may be much less relevant to comprehension -- unless they are guided by information that is available to everyone, especially evidence that a referent is attended by the speaker as well.

Indeed, existing evidence about the relationship between attention and reference comprehension is centered on questions about **shared** attention. Some studies have examined the effect of nonlinguistic cues to attention, like the eye gaze of one’s interlocutor. Baldwin (1991, 1993) found that toddlers are sensitive to the speaker’s gaze for interpreting the referent of novel words (but see Samuelson and Smith, 1998, for an alternate view). Hanna & Brennan (2008) found that adult listeners used speaker’s eye gaze to facilitate the comprehension of temporarily ambiguous expressions like *the blue circle with five dots on it*. Moreover, their participants utilized eye gaze flexibly, looking for the object in the opposite direction of the speaker’s gaze when their own display was arranged in reverse order of the speaker’s display. Hanna and Brennan concluded that eye gaze is not used automatically, suggesting that it is more of an endogenous than exogenous cue (but see Friesen & Kingstone, 1998). Likewise, Nappa & Arnold (under review) found that pronoun interpretation is influenced by both pointing and the speaker’s eye gaze. Reference comprehension is more generally facilitated if attention is jointly focused on a specific region of a display, limiting the domain of interpretation (Beun & Cremers, 1998, see also Brown-Schmidt, Campana, & Tanenhaus, 2005).

A related question is whether discourse information itself influences accessibility via attentional mechanisms. Foraker & McElree (2007) tested the proposal that listeners maintain linguistically prominent entities in focal attention, by examining discourses in which one entity appeared as the focus of a cleft (*It was the new foreman who unrolled the latest blueprint. He…*). The focus of clefts is known to make entities more accessible as pronoun referents (Almor, 1999). If they do so by causing the entity to be maintained in focal attention, Foraker & McElree predicted that pronouns referring to the focus of clefts would be read faster than those referring to nonclefted antecedents (*What the new foreman unrolled was the latest blueprint. He…).* In a speed-accuracy tradeoff paradigm, they found that clefting did not affect speed, but only the likelihood of successfully retrieving the antecedent for the pronoun. They concluded that clefting does not result in actively maintaining the referent representation in focal attention, but instead affected the strength with which the representation of the character was encoded.

In sum, there is clear evidence that reference comprehension is influenced by cues to shared attention, including eye gaze, pointing, and the shared discourse context. Yet even though these findings are consistent with the idea that attention guides accessibility (and/or strength of representation of the discourse entity), it is also possible that eye gaze and discourse context guide reference comprehension directly. The strongest test of the role of attention, therefore, is to examine how nonshared attention guides pronoun comprehension.

**Exogenous capture as a manipulation of attention**

We examined this question, drawing on the visual attention literature to find a non-discourse manipulation of the listeners’ attention. Several studies have shown that visual attention is automatically captured by sudden onset cues, for example a black square on a red background (Folk et al., 1992; Jonides, 1981; Hopfinger & Mangun, 1998; Müller & Rabbit, 1989). Evidence for attentional capture comes from tasks where the response to a visual target (e.g., two-choice letter discrimination) is faster when a capture cue preceded the target in the same location, as opposed to a different location.

This kind of capture cue is clearly not a part of the discourse itself. In situations like ours, where the cue is not predictive of the pronoun referent, there is no public evidence that it should be used for the purpose of calculating accessibility. Furthermore, such cues can often occur without explicit awareness on the part of the listener. If the listener is unaware that a sudden-onset cue has captured their attention, the experience is as if their attention landed on one character for idiosyncratic reasons, unrelated to the comprehension task. It is as if your attention is drawn to a metal cup because the sunlight suddenly glances off it in your direction. Language comprehension would be inefficient if such things influenced our understanding of pronouns.

On the other hand, people are not always adept at ignoring information that is egocentrically available when participating in a joint activity like conversation. For example, expressions like “the blue triangle” are interpreted by considering both what information is available to the speaker (i.e., what is in common ground), and what objects in one’s own ground are good matches for the input (Hanna, Tanenhaus, & Trueswell, 2003; Hanna & Tanenhaus, 2004; Keysar, Barr, Balin & Paek, 1998; Keysar, Barr, Balin, & Brauner 2000). Furthermore, any representations of joint attention must be processed by the listener’s internal cognitive system. Egocentrically attended information may therefore leak into calculations of shared attention. Such “leakage” effects are likely to be small, and may only hold when they are partially consistent with shared information. Nevertheless, effects of nonshared attention would provide the strongest evidence that pronoun comprehension is driven by attention per se, and not just discourse context.

Two experiments tested the role of nonshared attention on pronoun comprehension. In both experiments, we monitored listeners’ eye movements as they listened to a short story about two same-gendered characters, which included a pronoun that referred to one of them. Participants’ fixations following the pronoun provided evidence of their hypotheses about the likely referent of the pronoun (Arnold et al., 2000). A capture cue attracted attention to either the target (the pronoun referent) or the competitor (the other character).

Our use of a story context provided a fairly natural use of pronouns. It also meant that the effects of the capture cue were tested in the context of discourse cues, specifically the known preference to interpret a pronoun as co-referential with a first-mentioned/ subject referent, rather than a second-mentioned/ nonsubject referent (Arnold et al., 2000; Gordon et al., 1993). We predicted strong effects of the discourse context. The critical question was whether the capture cue would additionally increase the likelihood of considering the cued character as a referent for the pronoun.

In both experiments, we are particularly interested in the role that attention plays at the start of the discourse, as participants set up a representation of the story. In a story like *Doggy picked apples with Birdy…,* the story itself only provides partial information about who the discourse focus is. The first-mentioned/subject referent is generally considered more accessible (Arnold, 2010; Arnold, 2001; Jarvikivi et al., 2005; Kaiser & Trueswell, 2004), but this is only a partial constraint. That is, the linguistic cues to accessibility are ambiguous. In this situation, the listener’s attention at story onset may modulate the strength of the first-mention bias. This study is the first (to our knowledge) that examines the relationship between attention at the onset of a story and later pronoun comprehension.

**Experiment 1**

*Method*

*Participants*

A total of 86 psychology students at UNC Chapel Hill participated in exchange for course credit. Data from 13 participants were not analyzed at all, because the participants reported on a voluntary questionnaire that they had noticed the flashes (n = 11) or that they had attention deficit disorder (n=2). An additional 21 participants were excluded because of problems with calibration. The fixations for each participant were visualized using McMurray’s EyeAnal software, and those with excessive drifting were excluded. Participants were also excluded if more than 30% of their trials were unusable due to trackloss, or if the participant had not fixated either the target or competitor during the entire trial (i.e. they used peripheral vision). This left 52 participants in the analysis.

*Stimuli*

The basic task paradigm was adapted from that used by Arnold et al. (2000). Participants viewed a picture on a PC computer (see Figure 1) while listening to a spoken story. All experimental verbal stimuli contained two sentences, e.g. *Birdy picked apples with Doggy near the farmhouse. He was wearing a hat to protect himself from the sun.* The first sentence introduced two same-gender characters, either Doggy and Birdy (male), or Kitty and Bunny (female). Each character (Birdy, Doggy, Kitty, Bunny) was first-mentioned five times. The second sentence began with a pronoun that referred to one of the characters; this was the target word. The referent of the pronoun was either a) the first-mentioned (N1) or b) the second-mentioned (N2) character from sentence 1. The pronoun referent was manipulated by changing the picture so that the disambiguating characteristic (e.g., wearing a hat) coincided with the intended referent. The disambiguating word (e.g., *hat*) always occurred after the verb; for example, in this picture both characters are wearing something, but only one is wearing a hat. This created an ambiguous period during which the scene was consistent with both pronoun interpretations (*He was wearing a)*. The critical question was where participants would look in during the period of time immediately following the pronoun. During the ambiguous period, the proportion of looks to a character are assumed to reflect a consideration of that character as a referent, all other things being equal.

|  |
| --- |
| a) s2_n1 b) c) |
| Figure 1. Sample visual visual stimuli for Experiment 1, in a) first-mentioned pronoun condition, and b) second-mentioned pronoun condition. A sample first-mentioned capture cue is given in c). |

The experimental verbal stimuli used two verb types in the first sentence: 1) Joint-action, e.g. *Birdy picked apples with Doggy*, or *Kitty went sightseeing with Bunny*; and 2) Source-Goal, e.g. *Bunny showed a sculpture to Kitty*, or *Kitty floated over to Bunny*. Eight items were source-goal, and 12 were joint action. Verb type was controlled because it is known to influence the relative accessibility of its arguments; for example goal arguments are slightly more accessible than source arguments (Arnold, 2001; Rohde, Kehler, & Elman, 2007; Stevenson, Crawley, & Kleinman, 1994). However (see below), in this experiment it had little effect on our outcome measures.

The experimental visual stimuli contained two characters, one on each side of the screen. These were always in the same ports, equidistant from the center of the screen. A third object always appeared in the top center (e.g., the barn). These three objects were equidistant from the center and from each other. In some pictures there was also a fourth object in the center (e.g., the tree). The displayed picture took up the entire 1280 x 1024 resolution screen. The characters filled left and right ports that were 352 x 402 pixels in size.

Immediately preceding the story picture, a brief visual capture screen appeared for 200 msec. This screen pictured a 38 x 38-pixel black square in the location of one of the characters, as a manipulation of visual attention (cf. Gleitman et al., 2007). On filler trials the capture cue appeared in other locations.

|  |
| --- |
| Fig. 1 |
| Figure 2. Example timing of visual stimuli for Exp. 1 (N2 target, N1 cued condition). After participants initiated the trial, a capture cue appeared at the location of one character for 200 msec. The picture appeared for 200 msec before the story begins. (Story was heard only and not shown on screen) |

Figure 2 illustrates the time course of stimuli for each trial. Participants clicked a GO button to initiate the trial, which was followed by a white screen with a black square in the location of one of the two characters. After 200 msec the square disappeared and the picture appeared, and 200 msec later the story began to play. The goal of this experiment was to test whether attention would modulate pronoun processing, in a situation where it was clearly not a part of the discourse record. The strongest way to achieve this was to ensure that participants would not be aware of the presence of the capture cue. Pilot testing revealed that a 200-msec sudden onset cue was not noticed by most participants. Post-experimental questionnaires confirmed this, and anyone who reported any suggestion of awareness of the cue was excluded from analysis (N=11).

Each trial was followed by a screen asking participants to press a button to indicate if the story and picture matched. A following screen prompted participants to provide an explanation if they had said that the story and picture didn’t match on the previous screen. This allowed us to identify rejections of the story that revealed a strong preference to identify the pronoun with the other character. For example, in the second-mentioned condition of the example, the story said *He’s wearing a hat*, the picture showed Doggy in a hat. If the participant said *Birdy isn’t wearing a hat*, it was taken as evidence that they couldn’t accept Doggy as the pronoun referent.

*Design*

The experimental design was 2 (pronoun target N1 vs. N2) x 2 (target cued vs. target not cued). Several properties of the stimuli were balanced across the stimuli set, including verb type (joint action vs. source-goal), target location (left vs. right side of the screen), and N1 character (Doggy, Birdy, Kitty, Bunny). 20 experimental items were pseudo-randomized with 8 filler items in four lists, each of which had both a forwards and a backwards version. Of the 8 fillers, 6 had stories that were not intended to match the picture for reasons that had nothing to do with pronominal reference (e.g., *Doggy went boating one summer…* when the picture showed Doggy snorkeling). There were two practice items, one that matched and one that did not match. The practice and filler items had one, two, or three characters, but used different structures and had no ambiguous pronouns. There was an attentional capture cue on all fillers, but in different locations.

*Apparatus*

We monitored participants' eye movements with a head-mounted Eyelink II eyetracker as they viewed the display and listened to the story. Eye position was sampled at a rate of one datapoint every 4 msec (250 Hz), but the data were converted to a granularity of one data point for every 16 msec (62.5 Hz) prior to analysis to speed processing (McMurray, 2002). Corneal reflection monitoring was used when possible (n=37); the pupil-only monitoring mode was used if we could not achieve adequate calibration with the corneal reflection mode (n=27). We analyzed only one eye, using the Eyelink automatic procedure for choosing the eye with better calibration. Calibration always received the highest rating of “good” by the Eyelink program (max deviation: x = 26.1 pixels, y = 118.4 pixels), otherwise the participant was re-calibrated. At analysis, fixations were allowed to deviate by 100 pixels to account for possible calibration error or calibration drift over the course of the experiment. This did not result in any overlap in the critical character ports.

The visual and auditory stimuli were presented on a PC computer running the ExBuilder software (an in-house software created at the University of Rochester; Longhurst, 2006), running on a PC computer with a 19” monitor (resolution: 1280 x 1024 pixels, refresh rate: 75 Hz).

*Procedure*

We gave participants a general description of the task. Then participants were calibrated using Eyelink’s 9-point calibration and validation procedure. Participants sat at the display computer so they could comfortably reach the mouse; viewing distance was approximately 22-34 inches. After calibration, participants were introduced to the characters in the stories and asked to name sample pictures of each character. Participants then performed two practice trials and were given a chance to ask questions. Then the experimental session began. Participants clicked on a crosshair at the center of the screen to begin each trial. During each trial, participants viewed a picture and listened to the story. When each story finished playing, another screen appeared which asked participants to indicate whether the story matched the picture by clicking a YES or NO button. A following screen asked participants to explain any NO responses out loud to the experimenter.

A post-experiment questionnaire was used to assess participants’ knowledge of the purpose behind the experiment, and whether they were aware of the capture cue. Anyone who indicated awareness of the cue was not included in the analysis.

*Logic of analyses and statistical methods*

Of primary interest in this study are the mechanisms that occur during language comprehension, as listeners attempt to identify a referent for ambiguous pronouns. Our interest in listener’s **on-line** (i.e., momentary) biases stems from the assumption that they are related to the mechanism of arriving at a final interpretation. To this end, we first focus on evidence of participants’ gaze immediately after the pronoun, and consider how this relates to their final interpretation. As a secondary analysis, we also present the results of the off-line picture/story match task. In both cases, the purpose of this study is to examine the relationship between visual attention at the start of the story and later pronoun interpretation. We therefore also present descriptive analyses of eye gaze at the point of the visual capture manipulation.

*Eye gaze analysis procedure.*

*Dependent variable.* Eye movement data throughout the paper are presented in terms of **looks**, where a look is defined as a fixation grouped together with the prior saccade. This categorization is often used in language studies, which aim to measure the period of time during which listeners were directing their attention toward one referent and not another (e.g., Arnold, Hudson-Kam & Tanenhaus, 1997; McMurray, Tanenhaus, & Aslin, 2009; Sedivy, Tanenhaus, Chambers, & Carlson, 1999 and J. Sedivy, personal communication, March 2009; Trueswell, Sekerina, Hill, & Logrip, 1999 and J. Trueswell, personal communication; March 2009). Saccades are ballistic and therefore the onset of the saccade represents the decision to begin looking at that object, and this look continues until the launch of a new saccade. Saccades were identified using Eyelink’s on-line parser, which uses a velocity and acceleration-based detection algorithm. Using McMurray’s (2002) EyeLinkAnal program in Microsoft Access, fixations were grouped by area of interest (target, competitor, object 1, object 2, other). In all eye movement analyses, trials were excluded if there was more than 33% track loss during the ambiguous period following the pronoun or if the participant failed to fixate either target or competitor for the entire trial (2.3% of the data)

We are particularly interested in the proportion of time spent looking at the target and competitor characters immediately following the pronoun (300 msec following pronoun onset until 300 msec following the disambiguation point). We therefore calculated the empirical logit of the proportion target looks (Barr, 2009): log ((# samples target looks + .05)/# samples competitor looks + .05)).

*Analysis procedure and control variables.* The effects of experimental conditions were evaluated in a mixed-effects linear regression model, using SAS proc mixed, including random effects for both participants and items, as well as random slopes for manipulated variables with respect to both participants and items (Barr, Levy, Scheepers, & Tily, 2013). For each analysis reported, we first built a control model to assess the effects of potential control variables (verbtype; List (1,2,3,4); forward vs. backward order; first vs. second half; target location (left/right); participant gender (female/male), whether the participant was fixating the target at the onset of the pronoun (yes/no); whether the participant was fixating the competitor at the onset of the pronoun (yes/no). Any contributions that were significant at a level of t=1.5 or greater were retained for the final analysis.

The final analysis included the predictors of interest; for example, referent (first vs. second mention) and visual capture cue, which were also centered. Those control predictors that reached criterion were aslo included. Control variables were deemed important, even though they were also controlled by experiment design, because our final analyses used either a subset of the data or outcome variables as predictors (see below).

*Response analysis procedure.* The task that participants performed was to decide if the story matched the picture. The primary motivation for this task was to encourage time-locking between eye movements and the linguistic input (Arnold et al., 2000). An additional advantage is that the responses provided some information about listeners’ final interpretation of the story. On those trials where the participant said that the story and picture did not match, they also provided a verbal explanation of their response. This allowed us to identify items where the participant provided a reason that had something to do with the pronoun referent (e.g., *Birdy isn’t wearing a hat*). These were categorized as “Pronoun Mismatch” responses. All match responses and mismatches with non-pronoun related reasons (e.g., *It looks more like a barn, not a farmhouse*) were categorized as “Pronoun Match” responses. A potential concern with this coding method is that difficulty with processing the pronoun in some conditions (e.g., for N2 referents) might lead to more Mismatch responses, but if participants cannot articulate the problem with the pronoun, our measure would underestimate differences between conditions in responses. However, there was no evidence that this was a problem: participants were unlikely to provide non-pronoun-related explanations for both N1 (4%) and N2 (2%) conditions. Note that this response measure is not the same as asking them what they thought the pronoun referred to. Rather, it is a measure of whether they were willing to accept the picture as a valid representation of the story.

Because match responses were binary, we analyzed them in a multilevel logistic regression, using SAS proc glimmix (version 9.1) with a Penalized Quasi-Likelihood (PQL) estimator, with a binary distribution and a logit link. The same procedure for including control variables was used as for the eye movement analyses.

*Results*

The goal of this study is to test the hypothesis that listeners’ attention as a discourse develops can influence their biases as they seek out potential referents for a pronoun, where we manipulated attention with a brief visual capture. We therefore begin with a consideration of how the visual capture affected eye gaze at the onset of the trial. This motivates two analyses of the relationship between discourse-initial attention and pronoun resolution. The first analysis uses the subset of trials on which the capture cue was effective, and tests its effect on pronoun interpretation. The second analysis uses all data, but instead of the capture cue, uses fixations at the onset of the trial as a measure of the participant’s visual attention.

*Response to the capture cue: eye gaze at the start of the trial*

We first examined looks to the animate characters as the trial began, as a way to assess the efficacy of the capture cue, and as a way to investigate how eye gaze reacted to other characteristics of the task. We expected listeners to visually attend to the character who appeared at the location of the visual capture cue, which appeared for 200 msec immediately prior to the onset of the stimulus picture. We also expected attention at trial onset to be constrained by a bias toward the left-hand character (Dahan, Tanenhaus, & Salverda, 2007).

Figure 3 illustrates the average allocation of visual attention during the first sentence of the story. Looks are plotted as the proportion of looks to N1 out of both N1 and N2, and as a function of both screen location (left vs. right) and capture cue (N1 cued vs. not cued). As the figure shows, the flash (i.e., capture cue) attracted attention before picture onset. When N1 was cued, proportion N1 looks were very high. When N1 was not cued, looks to N1 were very low. Another way of saying this is that when N2 was cued, initial looks to N2 were high.

Figure 2 also shows that the effect of the capture cue was limited to the first second or so of the trial. Shortly after the story began at 400 msec after trial onset, looks were influenced by the verbal input, and showed a general bias toward the first-mentioned character. The overall N1 bias can be observed by examining looks above and below the 50% line. At trial onset (i.e., before the sentence begins), there was naturally no bias toward N1. Starting around 800 msec, most of the looks fall above the line, demonstrating a general N1 bias. By 1200 msec after trial onset, there was very little effect of cueing and location, and looks were primarily driven by the discourse itself. Figure 3 plots all looks from trial onset up until pronoun onset, which demonstrates that the effect of initial attention did not persist beyond the first sentence.

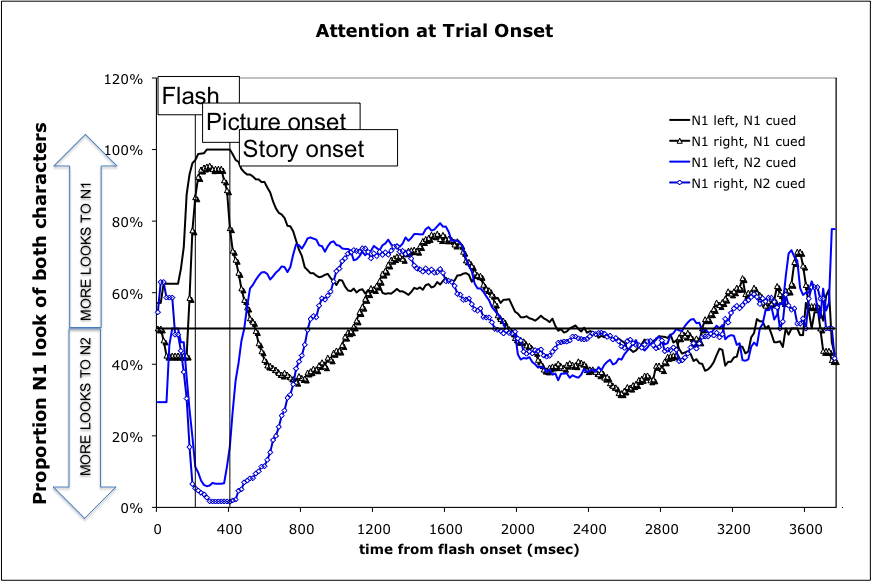


Figure 3. Proportion of looks to N1 (the first-mentioned character) out of all looks to the two characters (i.e., N1/(N1+N2)[[2]](#footnote-2) at each time point, as a function of the cueing and left/right position of N1. Note that when N1 was not cued, N2 was cued. Lines toward the top of the graph indicate greater looks to N1 than N2; lines toward the bottom of the graph indicate greater looks to N2 than N1.

Nevertheless, this figure illustrates that the capture cue did not attract attention categorically. That is, the capture cue was only partially effective. Our first analysis is therefore limited to trials on which the capture cue successfully manipulated trial-initial attention.

*Analysis 1: The effect of the capture cue on pronoun resolution*

To examine the effect of the effect of the capture cue, we limited the analysis to only those trials on which the capture cue was effective. As an index of effectiveness, we calculated the proportion of looks at the cued and uncued characters over the first second of the trial. When there were more looks to the cued character (62% of all trials), we were confident that the capture cue manipulated their attention.

Our primary dependent measure was eye gaze following the pronoun, from 300 msec after the pronoun until 300 msec after the onset of the disambiguating word. As predicted, there were more N1 looks than N2 looks overall (see Table 1). Critically, the capture cue also affected post-pronoun looks: there were more looks to N1 when it was cued, and more looks to N2 when it was cued.

Table 1. Experiment 1 results, Analysis 1. Average looks over the ambiguous region following the pronoun and percentage Match responses.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Target | Capture Cue | N1 looks | N2 looks | Target Empirical logit | %Match | N |
| N1 target | target cued | 0.56 | 0.18 | 1.86 | 98% | 173 |
|  | competitor cued | 0.51 | 0.30 | 0.93 | 98% | 142 |
| N2 target | target cued | 0.43 | 0.38 | -0.21 | 20% | 155 |
|  | competitor cued | 0.52 | 0.25 | -1.27 | 8% | 155 |

We assessed the reliability of these effects in a mixed-effects linear regression model, following the procedure outlined above. Table 2 lists the statistical summary for both critical predictors and control variables that were included in the final model. This model confirmed that post-pronominal looks to the target character were affected by both manipulations: 1) there were more looks to the target when it was N1 than N2, and 2) there were more looks to the target when it had been cued at trial onset.

Table 2. Parameter estimates, t-statistics and p-values for critical and control predictors in analysis 1. 625 observations included; data excluded for track loss and when trial-initial fixations did not match the capture cue. Models include random intercepts for participants and items, as well as slopes for pronoun referent by both participants and items, and capture cue by both participants and items.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | EYE MOVEMENTS | | | | RESPONSES | | | |
| Variable | Estimate  (error) | DF | t | p | Estimate  (error) | DF | t | p |
| CRITICAL PREDICTORS | | | | | | | | |
| N1 vs. N2 pronoun referent | 0.92(0.21) | 616 | 4.46 | <.0001 | 6.17(0.57) | 46.17 | 10.92 | <.0001 |
| Target vs. competitor cued | 0.32(0.12) | 616 | 2.73 | 0.006 | 0.9(0.36) | 619 | 2.49 | 0.013 |
| CONTROL PREDICTORS | | | | | | | | |
| List 1 | 0.25(0.16) | 616 | 1.55 | 0.121 | \_ | \_ | \_ | \_ |
| List 2 | 0.23(0.1) | 616 | 2.25 | 0.025 | \_ | \_ | \_ | \_ |
| List 3 | 0.1(0.06) | 616 | 1.69 | 0.092 | \_ | \_ | \_ | \_ |
| Forward list direction (vs. backward) | -0.27(0.1) | 616 | -2.64 | 0.009 | -0.44(0.38) | 54.31 | -1.17 | 0.245 |
| Target on left vs. right side | 0.26(0.2) | 616 | 1.31 | 0.19 | \_ | \_ | \_ | \_ |
| Target fixated at pronoun onset | 0.31(0.11) | 616 | 2.67 | 0.008 | 0.54(0.38) | 619 | 1.39 | 0.165 |
| Competitor fixated at pronoun onset | \_ | \_ | \_ | \_ | \_ | \_ | \_ | \_ |
| First half (vs. second half) | \_ | \_ | \_ | \_ | 0.77(0.34) | 619 | 2.24 | 0.025 |
| Female (vs. male) participant | \_ | \_ | \_ | \_ | \_ | \_ | \_ | \_ |
| Verb bias | \_ | \_ | \_ | \_ | \_ | \_ | \_ | \_ |

The next question was whether these effects also modulated participants’ off-line responses. Participants were highly likely to respond Match when the target was N1, regardless of whether the target was cued (98%) or not (98%). When the target was N2, the participants were more likely to respond Match when the target was cued (21%) than when it was not (8%).

We submitted these data to a mixed effects logistic regression, which confirmed that responses were affected by both target referent (more Match responses for N1 than N2), and by attentional cueing (more Match responses for cued targets than not cued targets). Although the means suggest that target referent interacted with attentional cueing, the interaction did not significantly contribute to the model, when it was included.

In summary, participants’ attention at the onset of each trial had an impact on both their online pronoun processing, and their final interpretation of the pronoun. Our first analysis demonstrated this by analyzing the effects of the capture cue, which drew participants’ attention to the cued character on 62% of the trials.

However, 38% of the data were not included in this analysis, because the capture cue was only a partial manipulation of attention. This raises questions about how pronoun comprehension is affected by attention in general, as opposed to the cueing manipulation per se. If the cueing manipulation represents general effects of attention during discourse comprehension, we might expect similar effects if we examine the participants’ attention at trial onset in all trials. This prediction is supported by the fact that participants were unaware of the cue, which means that from their perspective it was the same as if their attention at trial onset was driven by their own idiosyncratic preferences. Our next analysis therefore examined the hypothesis that any modulation of listeners’ visual attention at the onset of the trial should affect later pronoun resolution.

*Analysis 2: The effect of the trial-initial attention on pronoun resolution*

Our second analysis examined pronoun processing in the entire dataset, considering the effects of the participants’ actual visual attention during the first second of each trial. We did not restrict participants from making eye movements, which means that their looks were good indicators of their visual attention (e.g., Altmann & Kamide, 2007). As shown in Figure 3, looks during the first second were primarily driven by cueing and the left-side bias, while the linguistic input had relatively little input. We therefore chose to use the Target Advantage (i.e., target/(target + competitor)) during the first second as a metric of Trial-initial attention.[[3]](#footnote-3) This metric was then used as a predictor in subsequent analyses.

The effect of trial-initial attention is shown in Table 3 and Figure 4. For presentational purposes, we binned trials into either “High attention” (greater than 50% looks in the first second) or “Low attention” (fewer than 50% looks in the first second), although Trial-initial attention was entered into the statistical model as a gradient predictor. As Figure 4 shows, there were more looks to N1 than N2, for both N1 and N2 target conditions. This reflects the well-known N1 advantage. In addition, there was a small boost in post-pronominal fixations for characters that had been attended at trial onset.

Table 3. Experiment 1 results, Analysis 2. Average looks over the ambiguous region following the pronoun and percentage Match responses. Trials are categorized by visual fixations during the first second of the trial.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  | N1 looks | N2 looks | Target Empirical logit | %Match response | N |
| N1 target | target attended at trial onset | 0.56 | 0.18 | 1.83 | 97% | 281 |
|  | competitor attended at trial onset | 0.47 | 0.31 | 0.75 | 98% | 240 |
| N2 target | target cued | 0.42 | 0.36 | -0.25 | 21% | 271 |
|  | competitor cued | 0.52 | 0.27 | -1.14 | 11% | 252 |

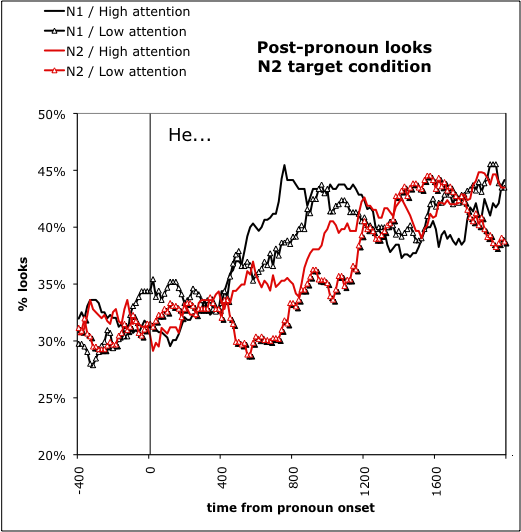
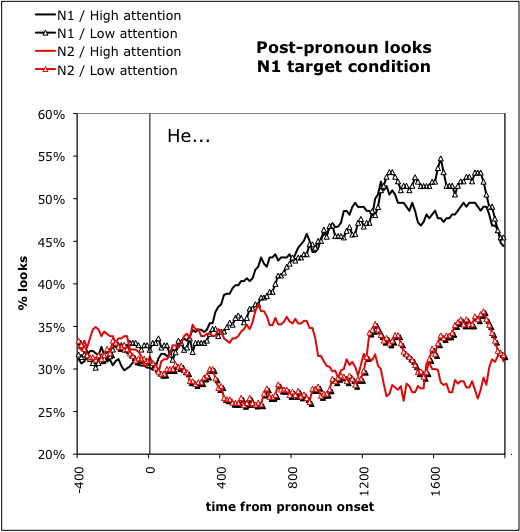


Figure 4. Post-pronominal looks to N1 and N2 as a function of Trial-initial attention and pronoun target condition.

Again, a multilevel mixed effects model was used to assess the effects of pronoun target (N1 and N2) and trial-initial attention on character looks, where the dependent variable was the empirical logit of target looks during the post-pronominal ambiguous region. For the model, trial-initial attention was included as a continuous predictor, using the proportion looks to the target out of all character looks. The flash manipulation was also included as a critical predictor.

As shown in Table 4, there was a significant effect of pronoun target, such that there were greater looks at N1 than N2 characters in the ambiguous region. There was also a significant effect of trial-initial attention, such that there were more post-pronominal target looks when the target had also been attended at trial onset. The flash manipulation itself did not contribute to this model.

Table 4. Parameter estimates, t-statistics and p-values for critical and control predictors in analysis 2. 1016 observations included; data excluded for track loss and when trial-initial fixations did not match the capture cue. Models include random intercepts for participants and items, as well as slopes for pronoun referent by both participants and items, and capture cue by both participants and items.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | EYE MOVEMENTS | | | | RESPONSES | | | |
| Variable | Estimate  (error) | DF | t | p | Estimate  (error) | DF | t | p |
| PREDICTORS OF INTEREST | | | | | | | | |
| N1 vs. N2 pronoun referent | 0.85(0.19) | 1004 | 4.55 | <.0001 | 6.17(0.53) | 29.54 | 11.58 | <.0001 |
| Target vs. competitor cued | 0.04(0.09) | 1004 | 0.44 | 0.664 | 0.16(0.27) | 1009 | 0.61 | 0.500 |
| Trial initial attention on target | 0.31(0.13) | 1004 | 2.31 | 0.021 | 0.73(0.43) | 932.7 | 1.69 | 0.091 |
| CONTROL PREDICTORS | | | | | | | | |
| List 1 | 0.21(0.14) | 1004 | 1.5 | 0.133 |  |  |  |  |
| List 2 | 0.17(0.08) | 1004 | 2.1 | 0.036 | - | - | - | - |
| List 3 | 0.08(0.05) | 1004 | 1.65 | 0.099 | - | - | - | - |
| Forward List direction | -0.27(0.09) | 1004 | -3.02 | 0.003 | - | - | - | - |
| Target on left vs. right side | 0.27(0.18) | 1004 | 1.53 | 0.126 | - | - | - | - |
| Target fixated at pronoun onset | 0.29(0.09) | 1004 | 3.16 | 0.002 | 0.62(0.31) | 1009 | 1.99 | 0.0466 |
| Competitor fixated at pronoun onset | 0.13(0.09) | 1004 | 1.49 | 0.1 | 0.39(0.3) | 1009 | 1.31 | 0.1894 |
| First vs. second half | - | - | - | - | 0.91(0.26) | 1009 | 3.46 | 0.0006 |
| Female vs. male participant | 0.27(0.12) | 1004 | 2.14 | 0.033 |  |  |  |  |
| Verb bias | - | - | - | - | - | - | - | - |

*Responses.* We also asked whether trial-initial attention affected responses in the full dataset. As shown in Table 4, there was a significant effect of target referent (N1 vs. N2), reflecting the finding that Match responses were more common for N1 than N2 referents. In addition, there was a small effect of trial-initial attention for responses in the N2 condition, which were more likely to be Match when N2 had been attended. This emerged as a marginally significant effect of trial-initial attention in the model. Even though trial-initial attention had a greater effect on N2 targets, the interaction term (target x trial-initial attention) was not significant when added to the model. Nonetheless, the marginal effect was clearly driven by the N2 target items. A separate analysis of just N2 target trials revealed a significant effect of trial-initial attention (t = 2.25; p = .025). Capture cue was not significant (p > .5).

Discussion

Experiment 1 demonstrated that participants’ attention at the onset of the trial is systematically related to two measures of pronoun resolution: 1) looks during the ambiguous region following the pronoun, and 2) responses to the offline picture verification task. These effects occurred despite the fact that participants were not aware of the visual capture cue that temporarily attracted their attention to one side of the screen. The same critical effects occurred in two different analyses. The first analyzed the effects of the capture cue, on those trials where it was effective as a manipulation of trial-initial attention. The second analysis included all data, and instead showed that looks at trial onset predicted looks after the pronoun and even marginally predicted off-line responses.

It is notable that the capture cue modulated fixations in the ambiguous region following the pronoun, despite reports in the literature that exogenous capture produces an inhibition of return (IOR) to the cued location after initial attention (Posner & Cohen, 1984). That is, if trial-initial attention is followed by subsequent inhibition, we might expect fewer, not more, looks to the attended character following the pronoun. The fact that we found facilitation from attention could stem from the fact that IOR effects tend to be fairly brief, in comparison with reference processing effects. Klein (2000) summarized eight IOR studies; at longer SOAs (comparable to the several second wait between capture cue and pronoun), responses to the cued location were about 10-50 ms faster than to the noncued location. By contrast, in this task we found differences among conditions spanning the entire ambiguous region (and beyond) – a period lasting on about 700 msec on average. Even if IOR effects do occur in this task, they are likely swamped by reference resolution-driven fixations.

At the same time, the effects of visual attention were secondary to the strong effects of order of mention. As expected, first-mentioned (N1) characters were preferred as the referents of the pronoun, both online and offline. The strength of order-of-mention as a discourse cue was especially apparent in the offline findings. This suggests that even though initial attention modulated on-line consideration of referents, public discourse cues had the strongest effects on final interpretation.

Nevertheless, it is notable that we found even a small effect of egocentric attention. Since none of the participants were aware of the capture cue, its effect is tantamount to any other pressure on the listener’s private attention. That is, there is no public information that the cued character should be attended – for all the listeners know, they just happened to attend to the cued character. This suggests that the critical effect of attention is not dependent on the capture cue.

Experiment 2 examines the extent of this effect, by asking whether the effects of trial-initial attention persist in the presence of a public, shared attentional capture cue. If the cue is more visually salient, will listeners view it as a part of the discourse record and ignore their own attentional biases?

**Experiment 2**

Experiment 2 pits the listeners’ own attentional biases at the onset of the story against a visually salient exogenous capture cue that occurred later in the story. This experiment permits us to dissociate the effects of attention at trial onset from the effects of shared attentional cues. The task was identical to that of Exp. 1, except the capture cue was a yellow halo around one character, which appeared in the middle of the first sentence. This manipulation had three key properties. First, it was not concurrent with trial onset, which meant that trial-initial attention could be assessed separately from the capture cue. Second, it occurred immediately preceding the mention of the second character in the first sentence of the story. This tests the alternate hypothesis that referential accessibility is affected by attention at the moment that the characters are mentioned. The second-mentioned character has the most ambiguous accessibility (i.e., it is somewhat accessible by virtue of having been mentioned, but not highly accessible like N1 is), and thus has the most potential to be affected by an attentional manipulation. Third, the capture cue was more salient than the cue in Experiment 1, and all participants were aware of it. This awareness was necessary, since any visual capture cue that disrupts the static scene would be noticed. This created a further contrast between the participant’s own attention at trial onset, and the salient, public visual capture cue.

*Method*

*Participants*

A total of 51 psychology students at UNC Chapel Hill participated for course credit. Data from there were not analyzed at all (1 reported having attention deficit disorder; and data for 2 participants data were lost due to technical problems. An additional 13 were excluded because of problems with track calibration, using the same criteria as for Exp. 1.

|  |
| --- |
| a.b.c. |
| Figure 5. Sample visual stimulus for Experiment 2. The main picture (a) is followed by a 200 ms presentation of either (b) or (c), followed again by (a). |

*Design, Materials, and Procedure*

The task, apparatus and procedure were identical to Experiment 1 except for the nature of the attention capture cue. A yellow halo around either N1 or N2 appeared for 200 msec, beginning 200 msec prior to the mention of the second character (see Figure 5). Since it takes about 200 msec to program and launch a saccade (Matin, Shao, & Boff, 1993), this should result in saccades to N2 around the time its name is mentioned. All other aspects of the design and procedure were the same, except that participants viewed the picture for 400 msec before the story began, not 200 msec.

*Results and Discussion*

As before, trials were excluded if there was more than 33% track loss during the post-pronominal ambiguous period, or if the participant failed to fixate either target or competitor for the entire trial (3.7% of all trials). We again categorized items by trial-initial attention during the first second of the trial, using the same procedure as in Exp. 1 (proportion target looks out of all target and competitor looks). Since the capture cue did not occur until between 1253 and 2244 msec into the trial, these early fixations were unaffected by the capture cue, and instead were driven by the participants’ own decisions about where to fixate. In this experiment participants spent more time fixating the left character in the first second (73%) than the right.

All participants reported after the experiment that they were aware of the visual capture. This meant that the results provided a good contrast between the effects of unconscious, internally-driven attention at the onset of the story, and a visual capture cue that was obvious yet not really a part of the story task itself.

Table 5. Experiment 2 results. Average looks over the ambiguous region following the pronoun and percentage Match responses. Trials are categorized by visual fixations during the first second of the trial.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  | N1 looks | N2 looks | Target Empirical logit | %Match | N |
| N1 target | target attended at trial onset | 0.52 | 0.22 | 0.61 | 0.93 | 180 |
|  | competitor attended at trial onset | 0.43 | 0.25 | 0.35 | 0.96 | 159 |
| N2 target | target attended at trial onset | 0.43 | 0.33 | -0.20 | 0.31 | 162 |
|  | competitor attended at trial onset | 0.43 | 0.27 | -0.33 | 0.23 | 173 |

*Eye movement analyses*

Again we analyzed the dependent variable of one or more looks to the target (or competitor) in the ambiguous region after the pronoun, i.e. from 300 after pronoun onset until 300 msec after the disambiguating word. Attention at trial onset was calculated in the same way as for Experiment 1.

As shown in Table 5 and Figure 6, target looks after the pronoun were more likely if the target had been attended at the onset of the trial. In addition, as expected, target looks after the pronoun were greater for N1 than N2 targets.

|  |
| --- |
|  |
| Figure 6. Experiment 2 Results. Percentage looks (average participant means) to the target and competitor over time, starting at the onset of the pronoun. Trials are divided by target referent (N1 vs. N2) and trial-initial attention (target fixated prior to competitor in first second of trial vs. not). The line indicates the average onset of the disambiguating word plus 300 msec. All data are included, regardless of fixation at pronoun onset. |

To assess the reliability of these patterns, these data were again submitted to linear mixed effects models, following the same procedure as for experiment 1. The control predictors were the same, except for the fact that target location (right/left) was excluded as a predictor. The reason for this is that trial-initial looks were primarily driven by screen location, where left-side targets were fixated more often.

As shown in Table 6, looks following the pronoun were influenced by both target referent (N1 vs. N2), and trial-initial attention. The capture cue manipulation had no effect on post-pronominal looks.

Table 6 Parameter estimates, t-statistics and p-values for critical and control predictors in the analysis for Experiment 2. 674 observations included; data excluded for track loss and when trial-initial fixations did not match the capture cue. Models include random intercepts for participants and items, as well as slopes for pronoun referent by both participants and items, and capture cue by both participants and items. Dashes indicate that the variable did not reach criterion in the control model and was not included in the final model.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | EYE MOVEMENTS | | | | RESPONSES | | | |
| Variable | Estimate  (error) | DF | t | p | Estimate  (error) | DF | t | p |
| PREDICTORS OF INTEREST | | | | | | | | |
| N1 vs. N2 pronoun referent | 0.71(0.12) | 664 | 5.69 | <.0001 | 4.38(0.54) | 38.79 | 8.05 | <.0001 |
| Target vs. competitor cued | -0.04(0.13) | 664 | -0.29 | 0.77 | 0.46(0.24) | 668 | 1.89 | 0.059 |
| Trial initial attention on target | 0.4(0.13) | 664 | 3.14 | 0.002 | 0.47(0.38) | 329.5 | 1.23 | 0.221 |
| CONTROL PREDICTORS | | | | | | | | |
| List 1 | 0.24(0.14) | 664 | 1.71 | 0.088 | - | - | - | - |
| List 2 | -0.13(0.16) | 664 | -0.85 | 0.398 | - | - | - | - |
| List 3 | 0.09(0.14) | 664 | 0.61 | 0.54 | - | - | - | - |
| Forward List direction | 0.2(0.1) | 664 | 2.01 | 0.044 | - | - | - | - |
| Target fixated at pronoun onset | 0.2(0.11) | 664 | 1.76 | 0.078 | 0.13(0.27) | 668 | 0.46 | 0.644 |
| Competitor fixated at pronoun onset | 0.22(0.11) | 664 | 1.97 | 0.05 | - | - | - | - |
| First vs. second half | - | - | - | - | 0.39(0.24) | 668 | 1.58 | 0.114 |
| Female vs. male participant | - | - | - | - | - | - | - | - |
| Verb bias | - | - | - | - | - | - | - | - |

*Response Data Analysis*

As shown in Table 5, participants were more likely to provide Match responses for N1 than N2. They were also numerically more likely to respond Match when N2 had been attended at trial onset than when it had not. However, the effect of trial-initial attention was not significant in the model (See table 6). Instead, there was a marginal effect of the capture cue on responses. Match responses were higher for N2 targets that were cued (31%) than those that were not cued (24%), while Match responses were high for N1 targets overall (cued: 95%; uncued: 93%).

*Discussion*

Experiment 2 provided further evidence that on-line pronoun resolution is influenced by the participants’ attention at trial onset. This effect emerged in a task in which visual attention was cued publically, with an obvious yellow flash cue in the middle of the story. Despite the salience of the visual cue, trial-initial attention had a greater effect than the capture cue on where participants looked following the pronoun. These looks are taken to indicate the participants’ initial consideration of referents for the pronoun, which suggests that early attention on one character is important to later pronoun processing. As in Experiment 1, trial-initial attention was an egocentric experience: participants were aware of the later flash, but initial attention was guided by stimuli characteristics that are not usually considered a part of the discourse record, such as being on the left side of the screen.

In contrast with Experiment 1, however, in this experiment the participants’ offline responses were somewhat guided by the capture cue, and not by trial-initial attention. This is likely to result from the salient nature of the capture cue in this experiment. Participants were all aware of the yellow halo, and they may have taken it as an intended signal about the importance of the N2 character to the story. Other studies have shown that listeners can use social cues, like pointing, to help identify referents for ambiguous pronouns (Nappa & Arnold, under review; Goodrich & Hudson Kam, 2009; Goodrich Smith & Hudson Kam, 2012).

Nevertheless, the attentional effects in this experiment were again secondary to the effects of the discourse. The predominant finding was that listeners preferred to link pronouns with N1 targets. This replicates the well-established finding that entities in subject or first-mentioned position tend to be salient, and preferred as pronoun targets (Arnold et al., 2001; Gernsbacher & Hargreaves, 1988; Gernsbacher et al., 1989; Jarvikivi et al., 2005; Kaiser & Trueswell, 2004).

**General Discussion**

Two experiments demonstrated that when listeners consider potential referents for a pronoun, they are initially biased toward the character that they attended at the beginning of the story. In Experiment 1, we manipulated their attention with a brief visual capture cue before the story began. Even though participants weren’t aware of the visual capture, the resulting attention on one character modulated their on-line biases as they considered potential referents for the pronoun. It turned out that the manipulation was not necessary to achieve this effect, in that trial-initial attention predicted on-line biases whether the attended character had been cued or not (see Gleitman et al., 2007, for a similar effect on production). Experiment 2 confirmed the importance of trial-initial attention, and found that a trial-medial capture cue did not modulate post-pronominal fixations.

While trial-initial attention had a consistent effect on participants’ on-line biases as they encountered the pronoun, it had a less consistent effect on their final responses about whether the story and picture matched. In experiment 1, early attention led to a marginal increase in Match responses, especially for N2 targets. In experiment 2, early attention had no effect on responses. This suggests that the effect is fragile. When participants were aware of the capture cue, they may have interpreted it as an intentional indication of importance to the discourse. This led to a small increase in Match responses, and overrode any effect of trial-initial attention.

*The role of attention in referential accessibility*

These data provide clear evidence that attention at the onset of a story increases the likelihood of considering the attended character, even briefly, when an ambiguous referential expression is encountered. At the broadest level, these findings support claims from the literature that attention is related to referential accessibility. By measuring visual attention, we have shown that discourse comprehension is influenced by attentional mechanisms. Although this hypothesis is well accepted in the literature, it has never been tested explicitly. In this study we focused on egocentric attention, which provided the strongest test that pronoun comprehension is affected by attention itself, and not the discourse or social contexts that modulate attention. Thus, this is the first solid evidence that attention is related to pronoun comprehension.

Moreover, our findings are informative about the nature of the relationship between attention and referential accessibility. Some theories of discourse processing treat focus of attention as a discourse category, rather than a psychological process, labeling it the discourse “focus”, or “center” (e.g., Grosz, Joshi, & Weinstein, 1994; Grosz & Sidner, 1986; Stevenson et al., 2000). While discourse structure clearly does influence accessibility, this study provides some of the first independently motivated evidence that psychological attention does too.

*The timecourse of attentional effects during reference processing*

The results of both experiments suggest that the listener’s direction of attention can modulate accessibility, but not necessarily at all points during the story. Rather, our attentional effects stemmed from attention at the onset of the trial. In Experiment 1, we manipulated trial-initial attention, and found that it affected both on- and off-line processing of the pronoun. In Experiment 2, we manipulated attention later in the trial, and found that it had no effect on participants’ on-line consideration of potential referents. Instead, their initial at trial onset again modulated looks following the pronoun.

This demonstrates that what is important is attention as the story representation is being built, and not attention when the pronoun is encountered. This finding is consistent with results from an experiment with children, in which that 4-5 year old children tended to assign ambiguous pronouns to referents they had been fixating at pronoun onset (Arnold, Brown-Schmidt & Trueswell, 2007).

The importance of trial-initial attention has implications for the possible mechanism by which attention affects accessibility. In particular, it suggests that attention is important in the establishment of a discourse representation, in which some characters are represented with greater accessibility. We know that discourse processing involves the construction of non-linguistic representations of discourse entities and events (e.g., van Dijk & Kintsch, 1983; Kintsch, 1988; Johnson-Laird, 1983; Bransford, Barclay & Franks, 1972, Bower & Morrow, 1990; Sanford & Garrod 1981; Zwaan & Radvansky 1998). These representations may be more or less activated, where activation is gradient and influenced by multiple sources of information (Ariel, 1990, 2001). We hypothesize that trial-initial attention increases the activation of the representations of discourse entities, making them more available when the pronoun is encountered. It is at this point that the listener is instantiating a discourse representation, which likely involves choices about their perspective on the discourse. The relatively more attended entities are represented with more activation. Under this view, discourse cues should also influence the activation of each representation. While we do not have specific evidence that attentional mechanisms also mediate the influence of discourse information, it is possible that they do.

We conjecture that trial-initial attention was relevant in these experiments precisely because it precedes the linguistic information itself. The linguistic first-mentioned bias is inherently probabilistic (e.g., Arnold et al., 2000) – while pronouns do tend to refer to first-mentioned characters, they can also refer to second-mentioned characters. At the onset of the story, listeners need to establish their perspective on the story. Their private attentional shifts may influence the way they interpret subsequent, probabilistic linguistic input, either strengthening or weakening the first-mentioned bias.

Another way of saying this is that we are not observing the effect of fleeting visual attention. Although we used exogenous (stimulus-driven) manipulations of attention, the effects we observed are likely to have been the result of an interaction with endogenous (goal-directed) processing. The capture cue in Experiment 1 attracted attention one character, immediately followed by the presentation of one of the characters in the same location. Once fixating there, participants may have decided to examine the character and think about its role in the story. This is consistent with the idea that endogenous (task-related) goals mitigate the effect of exogenous cues (Folk et al., 1992).

The importance of trial-initial attention is also apparent in a comparison of our findings with those of Nappa & Arnold (under review, Exp. 2). They had participants watch videos of a speaker telling a story about two co-present puppets. In one of the experiments, a capture cue (a black square) appeared at the location of one puppet. But in contrast with the current experiments, the capture cue occurred at the moment the pronoun was uttered. Even though the capture cue increased judgments about character location, it did not influence either judgments about the pronoun referent or speed of response. By contrast, the only relevant attentional cues at pronoun onset were social ones: pointing, gazing, or a novel condition in which the speaker claimed that she controlled the black squares on screen. These findings are consistent with the idea that the role of egocentric attention only has it effect as character representations are being formed, early in the story. Once these representations are established, later modulations of egocentric attention are irrelevant.

The task-relevance of the cued characters in our experiment may also explain another apparent paradox: visual attention is often attracted to new information, whereas linguistic discourse cues typically attract attention to old or topical information – yet both facilitated on-line pronoun resolution. This may have occurred because the critical moment of attention occurred immediately preceding the onset of the story. All task information at this point was new. But this was also the point at which listeners began to assess exactly how central each mentioned entity was to the story. Early attention on one character may have increased the expectancy of hearing a later reference to that character, which would have then facilitated subsequent reference comprehension (Arnold, 1998; 2008).

Our interpretation of the results is consistent with a characterization of accessibility as a gradient property of entity representations. This is consistent with Foraker & McElree’s (2007) claim that discourse accessibility does not involve putting a single entity into a privileged, focal category. In our experiments, attention only subtly modulated referential accessibility, showing no evidence that attention led to the privileged focusing of the attended character. It also seems unlikely that participants were aware of their visual attention at trial onset, suggesting their attention was not being actively directed or maintained.

*Shared vs. egocentric attention.*

A common view is that referential accessibility specifically emerges from assumptions about the attention or knowledge of one’s discourse partners (e.g., Ariel, 1990; Brennan, 1995; Chafe, 1994; Gundel et al., 1993). Here we have demonstrated that private attentional modulations also influence on-line biases during pronoun comprehension. The capture cue influenced processing in Experiment 1 even though participants were not aware of it. Moreover, we observed a more general effect of participants’ own decisions about where to attend in the first second of the trial, which can hardly be taken as evidence of shared attention.

This provides striking evidence that the effects of attention on referential accessibility are not limited to situations where attention is clearly shared. However, we want to emphasize that the importance of this finding is that it provides clear evidence for the role of attention. Any demonstration that pronoun interpretation is driven by public cues to attention can be attributed to either effects of attention, or effects of the public cues themselves. By contrast, evidence of egocentric attention provides a clean demonstration of attentional processes.

By contrast, we do not take these results as evidence that pronoun comprehension is primarily the result of egocentric representations of accessibility. While we found significant effects of trial-initial attention on pronoun comprehension, they emerged most strongly in our measures of on-line consideration of potential referents, i.e. in the eye movement analyses. Responses were overwhelmingly driven by linguistic status: pronouns referring to first-mentioned characters were accepted far more often than pronouns referring to second-mentioned characters. This is consistent with claims that reference should be interpreted with respect to shared information (e.g., Ariel, 1990; Gundel et al., 1993; Clark, 1996; Chafe, 1994).

Nevertheless, these findings suggest that listeners may not always keep track of the source of their attentional modulations. Whether they attend to a character because it is important to the story, or because it occurred on the left side of the screen, this attention still contributes to later pronoun processing. This is consistent with mounting evidence the mechanisms of language processing do not categorically ignore entities that are unknown to their speaker when identifying referents for nominal expressions (Barr & Keysar, 2006; Hanna, Trueswell, & Tanenhaus, 2003; Hanna & Tanenhaus, 2004; Keysar, Barr, Balin & Brauner, 2000).

*Conclusions*

In sum, we have provided evidence that private shifts of attention influence both listener’s transitory biases during pronoun resolution, and to a lesser extent, listener’s final interpretation of the pronoun. This study provides the first evidence about the relationship between attention at trial onset, and later reference resolution, in a visual-world eyetracking study. It also supports widespread claims that pronoun comprehension is driven by modulations in the attention of discourse participants, which make some referents more accessible than others.

**Appendix**

Verbal stimuli for Experiments 1 and 2

Source-Goal items:

Birdy brought some mail to Doggy during a big rainstorm. He was carrying an umbrella, which was a lucky thing.

Bunny showed a sculpture to Kitty outside the art museum. She had brought a camera to take pictures of the artwork.

Doggy gave some batting tips to Birdy on a sunny day at the park. He held the bat carefully to ensure a good grip.

Doggy read a story to Birdy near a tree outside. He sat on a log because the ground was wet.

Kitty floated over to Bunny at the pool party. She had a bag of chips for everyone to eat.

Kitty sang a song to Bunny at the school car wash. She scrubbed the window carefully so it would sparkle in the sun.

Kitty taught a song to Bunny at the neighborhood cookout. She sang really badly and ruined the song.

Kitty taught the hula to Bunny near some palm trees. She took a sip of her soda because it was a hot day.

Joint-Action items:

Birdy picked apples with Doggy near the farmhouse. He was wearing a hat to protect himself from the sun.

Birdy played board games with Doggy during the snowstorm. He picked up the dice because it was his turn to roll.

Birdy walked around the amusment park with Doggy in the rain. He got an ice cream cone even though it was cold and rainy.

Birdy went rollerskating with Doggy at the playground. He brought a frisbee so they could play with it later.

Bunny baked some cookies with Kitty with a brand new recipe. She took out a carton of milk to go with the cookies.

Bunny drew pictures with Kitty beneath the rainbow. She used a paintbrush to capture the vibrant colors.

Bunny threw a party with Kitty with a big piñata. She took a piece of cake and accidentally knocked the rest of it off the table.

Bunny went shopping with Kitty at the new mall. She wanted to buy a new watch because her old one was broken.

Doggy ate a snack with Birdy at the fruit stand He chose an apple because it looked really fresh.

Doggy made soup with Birdy in a big pot on the stove. He picked up the spoon to taste the soup.

Doggy went kite flying with Birdy near a bench in the park. He stood on a rock so that his kite would fly highest.

Kitty went sightseeing with Bunny at the old Spanish castle. She had brought a guide book for more information.

Author Note

We are grateful to Florian Jaeger, Daniel Serrano, Daniel Bauer, and Chris Weisen for discussion of statistical methods. Many thanks also to Cathleen Sparks, Glenn Kern, and Jordan Todd for their help preparing stimuli and collecting data. This research was partially supported by NIH grant R01 HD-41522 and partially supported by NSF grant BCS-0745627.

References

Almor, A. (1999). Noun-phrase anaphora and focus: The informational load hypothesis. *Psychological Review*, *106*(4), 748-765.

Altmann, G.T.M., & Kamide, Y. (2007). The real-time mediation of visual attenti on by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *Journal of Memory and Language, 57*, 502-518.

Ariel, M. (1990). *Accessing noun-phrase antecedents.* London: Routledge.

Ariel, M. (2001). Accessibility theory: An overview. In T. Sanders, J. Schilperoord, & W. Spooren (Eds.), Text representation: Linguistic and psycholinguistic aspects (pp. 29-87). Amsterdam: Benjamins.

Arnold, J. E. (1998). Reference form and discourse patterns (Doctoral dissertation, Stanford University, 1998). *Dissertation Abstracts International, 59,* 2950.

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

Arnold, J. E. (2008). Reference Production: Production-internal and Addressee-oriented Processes. *Language and Cognitive Processes*, 23, 495-527.

Arnold, J. E. 2010. How speakers refer: The role of accessibility. Language and Linguistic Compass 4. 187–203.

Arnold, J. E., Brown-Schmidt, S., & Trueswell, J. C. (2007). Children’s use of gender and order-of-mention during pronoun comprehension. *Language and Cognitive Processes, 7-8,* 1-40.

Arnold, J. E., Eisenband, J. G., Brown-Schmidt, S., & Trueswell, J. C. (2000). The rapid use of gender information: Evidence of the time course of pronoun resolution from eye tracking. *Cognition, 76,* B13–B26.

Arnold, J. E., Wasow, T., Losongco, T., & Ginstrom, R. (2000). Heaviness vs. Newness: the effects of structural complexity and discourse status on constituent ordering*. Language, 76*, 28-55.

Baldwin, D. A. (1991). Infants' contribution to the achievement of joint reference. *Child Development, 62*, 875-890.

Baldwin, D. A. (1993). Infants' ability to consult the speaker for clues to word reference. *Journal of Child Language, 20*, 395-418.

Bard, E. G., Anderson, A. H., & Sotillo C. (2000). Controlling the intelligibility of referring expressions in dialogue. *Journal of Memory and Language, 42*, 1-22.

Barr, D. J. (2009). Analyzing 'visual world' eyetracking data using multilevel logistic regression. *Journal of Memory and Language,* 457-454.

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

Barr, D. J., Levy, R., Scheepers, C. & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68(3):255-278.

Berger, A., Henik, A., & Rafal, R. (2005). Competition between endogenous and exogenous orienting of visual attention. *Journal of Experimental Psychology: General, 134*, 207-221.

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

Bock, J. K. (1982). Toward a cognitive psychology of syntax: Information processing contributions to sentence formulation. *Psychological Review, 89,* 1–47.

Bock, J. K. (1986). Meaning, sound, and syntax: Lexical priming in sentence production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12,* 575–586.

Bock, J. K., & Irwin, D. E. (1980). Syntactic effects of information availability in sentence production. *Journal of Verbal Learning and Verbal Behavior, 19,* 467–484.

Bock, K., Irwin, D. E., & Davidson, D. J. (2004). Putting first things first. In J. M. Henderson, & F Ferreira. (Eds.), *The interface of language, vision, and action: Eye movements and the visual world* (pp 249 – 278).New York: Psychology Press.

Bower, G. H., & Morrow, D. G. (1990). Mental models in narrative comprehension. *Science, 247*(4938), 44-48.

Bransford, J. D., Barclay, J. R., & Franks, J. J. (1972). Sentence memory: A constructive versus interpretive approach. *Cognitive Psychology, 3*(2), 193-209.

Brennan, S. E. (1995). Centering attention in discourse. *Language and Cognitive Processes, 102,* 137–167.

Brockmole, J. R., & Boot, W. R. (2009). Should I stay or should I go? Attentional disengatement from visually unique and unexpected items at fixation. *Journal of Experimental Psychology: Human Perception and Performance*, 35, 808-815.

Brown-Schmidt, S., Campana, E., & Tanenhaus, M. K. (2005). Real-time reference resolution by naive participants during a task-based unscripted conversation. In J. C. Trueswell, & M. K. Tanenhaus (Eds.), *Approaches to studying world-situated language use: Bridging the language as product and language as action traditions* (pp. 153–172). Cambridge, MA: The MIT Press.

Chafe, W. L. (1976). Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In C. N. Li (Ed.), *Subject and topic* (pp. 25-56). New York: Academic Press.

Chafe, W. L. (1994). *Discourse, consciousness, and time.* Chicago: Chicago University Press.

Chanon, V. W., & Hopfinger, J. B. (2008). Memory’s grip on attention: The influence of item memory on the allocation of attention*. Visual Cognition 16*, 325-340.

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

Clark, H. H., & Sengul, C. J. (1979). In search of referents for nouns and pronouns. *Memory & Cognition, 7*, 35-41.

Clark, H.H., & Marshall, C.R. (1981). Definite reference and mutual knowledge. In A. K. Joshi, B. L. Webber, & I. A. Sag (Eds.), *Elements of discourse understanding* (pp 10-63). Cambridge: Cambridge University Press.

Cowles, W.; Walenski, M., & Kleunder, R. (2007). Linguistic and cognitive prominence in anaphor resolution: topic, contrastive focus and pronouns. *Topoi* 26, 3–18, DOI 10.1007/s11245-006.

Dahan, D., Tanenhaus, M. K., & Chambers, C. G. (2002). Accent and reference resolution in spoken language comprehension. *Journal of Memory and Language, 47,* 292-314.

Dahan, D., Tanenhaus, M. K., Salverda, A. P. (2007). How visual information influences phonetically-driven saccades to pictures: effects of preview and position in display. In R. van Gompel, M. Fischer, W. Murray, & R. Hill (Eds.), Eye-movements: a window on mind and brain. Oxford, England: Elsevier.

De Graef, P., Christiaens, D., & d’Ydewalle, G. (1990). Perceptual effects of scene context on object identification. *Psychological Research, 52,* 317-329.

Ferreira, V. S. (2003). The persistence of optional complementizer production: Why saying ‘that’ is not saying ‘that’ at all. *Journal of Memory and Language, 48,* 379–398.

Ferreira, V. S., & Dell, G. S. (2000). Effect of ambiguity and lexical availability on syntactic and lexical production. *Cognitive Psychology, 40,* 296–340.

Folk, C. L., Remington, R.W., & Johnston, J.C. (1992). Involuntary covert orienting is contingent on attentional control settings. *Journal of Experimental Psychology: Human Perception and Performance, 18,* 1030-1044.

Foraker, S., & McElree, B. (2007). The role of prominence in pronoun resolution: Active vs. passive representations. *Journal of Memory and Language, 56,* 357-383.

Friesen, C. K., & Kingstone, A. (1998). The eyes have it! Reflexive orienting is triggered by nonpredictive gaze. *Psychonomic Bulletin & Review, 5,* 490–495.

Garnham, A., Traxler, M., Oakhill, J., & Gernsbacher, M. A. (1996). The locus of implicit causality effects in comprehension. *Journal of Memory and Language, 35,* 517-543.

Gernsbacher, M. A., & Hargreaves, D. J. (1988). Accessing sentence participants: the advantage of first mention. *Journal of Memory and Language*, 27, 699–717.

Gernsbacher, M. A., Hargreaves, D. J., & Beeman, M. (1989). Building and accessing clausal representations: the advantage of first mention versus the advantage of clause recency. *Journal of Memory and Language*, 28, 735–755.

Givón, T. (1983). Topic continuity in discourse: An introduction. In T. Givón (Ed.), *Topic continuity in discourse: A quantitative cross-language study* (pp. 1–42). Amsterdam: John Benjamins Publishing.

Gleitman, L. R., January, D., Nappa, R., & Trueswell, J. C. (2007). On the *give* and *take* between event apprehension and utterance formulation. *Journal of Memory and Language, 57,* 544-569.

Goodrich Smith, W., & Hudson Kam, C.L. (2012). Pointing to ‘her’: The effect of co-speech gesture on pronoun resolution. Language and Cognition, 4, 75-98.

Goodrich, W., & Hudson Kam, C.L. (2009). Co-speech gesture as input in verb learning. Developmental Science, 12, 81-87.

Gordon, P. C., Grosz, B. J., & Gilliom, L.A. (1993). Pronouns, names, and the centering of attention in discourse. *Cognitive Science 17*, 311-347.

Grosz, B. J., Joshi, A. K., & Weinstein, S. (1995). Centering: A framework for modeling the local discourse. *Computational Linguistics, 21*, 203-225.

Grosz, B., & Sidner, C. (1986). Attention, intentions, and the structure of discourse. *Computational Linguistics, 12,* 175–204.

Gundel, J. K., Hedberg, N., & Zacharski, R. (1993). Cognitive status and the form of referring expressions in discourse. *Language,* *69,* 274-307.

Hanna, J. E., & Brennan, S. E. (2008). Speakers' eyegaze disambiguates referring expressions early during face-to-face conversation. *Journal of Memory and Language, 57,* 596-619.

Hanna, J. E., & Tanenhaus, M. K. (2004). Pragmatic effects on reference resolution in a collaborative task: Evidence from eye movements. *Cognitive Science: A Multidisciplinary Journal, 28*(1), 105-115.

Hanna, J. E., Tanenhaus, M. K., & Trueswell, J. C. (2003). The effects of common ground and perspective on domains of referential interpretation. *Journal of Memory and Language, 49*(1), 43-61.

Hawkins, J. A. (1994*). A performance theory of order and constituency*. Cambridge: Cambridge University Press.

Henderson, J. M., Weeks, P. A., & Hollingworth, A. (1999). The effects of semantic consistency on eye movements during complex scene viewing. *Journal of Experimental Psychology: Human Perception and Performance, 25,* 210-228.

Hopfinger, J. B., & Mangun, G. R. (1998). Reflexive attention modulates processing of visual stimuli in human extrastriate. *Psychological Science, 9,* 441-447.

Hopfinger, J.B., & Ries, A.J. (2005). Automatic vs. contingent mechanisms of sensory-driven neural biasing and reflexive attention. *Journal of Cognitive Neuroscience, 17,* 1341-1352.

Järvikivi, J., van Gompel, R. P. G., & Hyönä, J. (2005). Ambiguous pronoun resolution: Contrasting the first-mention and subject-preference accounts. Psychological Science, 16(4), 260-264.

Johnson-Laird, P. N. (1983). *Mental models: Toward a cognitive science of language, inference and consciousness.* Cambridge, MA: Harvard University Press.

Jonides, J. (1981). Voluntary vs. automatic control over the mind’s eye’s movement. In J. B. Long, & A. D. Baddeley (Eds.), *Attention and performance IX* (pp. 187-203). Hillsdale, NJ: Erlbaum.

Jonides, J. & Yantis, S. (1988). Uniqueness of abrupt stimulus onset in capturing attention. *Perception & Psychophysics, 43*, 346-354.

Kaiser E., & Trueswell J. C. (2008). Interpreting pronouns and demonstratives in Finnish: evidence for a form-specific approach to reference resolution. *Language and Cognitive Processes*, 23:709–748.

Kaiser, E., & Trueswell, J.C. (2004). The role of discourse context in the processing of a flexible word-order language. *Cognition, 94,* 113-147.

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

Keysar, B., Barr, D.J., Balin, J.A., & Paek, T.S. (1998). Definite reference and mutual knowledge: Process models of common ground in comprehension. *Journal of Memory and Language*, *39*, 1–20.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A constructive-integration model. *Psychological Review, 95,* 163-182.

Klein, R. M. (2000). Inhibition of return. *Trends in Cognitive Sciences, 4,* 138-147.

Koorneef, A. W., & van Berkum, J. J. A. (2006). On the use of verb-based implicit causality in sentence comprehension: Evidence from self-paced reading and eye tracking. *Journal of Memory and Language, 54,* 445-465

Levelt, W. J. M. (1989). *Speaking.* Cambridge: MIT Press.

Loftus, G. R., & Mackworth, N. H. (1978). Cognitive determinants of fixation location during picture viewing. *Journal of Experimental Psychology: Human Perception and Performance, 4,* 565-572.

Longhurst, E. (2006). ExBuilder. [Computer program]

Marslen-Wilson, W.D., Tyler, L.K., & Koster, J. (1993). Integrative processes in utterance resolution. Journal of Memory and Language, 32, 647–666.

Matin, E., Shao, K. & Boff, K. (1993). Saccadic overhead: Information-processing time with and without saccades. *Perception and Psychophysics, 53*, 372-380.

McKoon, G., Greene, S.B., & Ratcliff, R. (1993). Discourse models, pronoun resolution, and the implicit causality of verbs. Journal of Experimental Psychology: Learning,Memory and Cognition, 18, 1040–1052.

McMurray, B. (2002). EyelinkAnal, version 1.5. Analysis scripts written in Microsoft Access.

McMurray, B., Tanenhaus, M.K., and Aslin, R.N. (2009) Within-category VOT affects recovery from "lexical" garden paths: Evidence against phoneme-level inhibition*. Journal of Memory and Language, 60*, 65-91

Morrow, D. G., Bower, G.H., & Greenspan, S. L. (1987). Accessibility and situation models in narrative comprehension. *Journal of Memory and Language, 26*, 165-187.

Morrow, D. G., Bower, G.H., & Greenspan, S. L. (1989). Updating situation models during narrative comprehension. *Journal of Memory and Language, 28*, 292-312.

Müller, H.J., & Rabbit, P.M.A. (1989). Reflexive and voluntary orienting of visual attention: Time course of activation and resistance to interruption. *Journal of Experimental Psychology: Human Perception and Performance, 15*, 215-330.

Nappa, R., & Arnold, J. E. (2013, under review). Paying Attention to Intention: Effects of Intention (But Not Egocentric Attention) on Pronoun Resolution. Manuscript; University of North Carolina at Chapel Hill.

Posner, M.I. & Cohen, Y. (1984). Components of visual orienting. In H. Bouma, & D. Bouwhuis (Eds.), *Attention and Performance* (Vol X, pp. 531-556). London, England: Erlbaum.

Remington, R. W., Johnston, J. C., & Yantis, S. (1992). Involuntary attentional capture by abrupt onsets. *Perception and Psychophysics, 51*, 279-290.

Rohde, H., Kehler, A., & Elman, J. (1997). Pronoun interpretation as a side effect of discourse coherence. *Proceedings of the 29th Annual Conference of the Cognitive Science Society*, pp. 617-622.

Salverda, A.P., & Altmann, G.T.M. (2005, September). *Cross-talk between language and vision: Interference of visually-cued eye movements by spoken language.* Poster presented at the *AMLaP Conference*, Gent (Belgium).

Samuelson, L. K., & Smith, L. B. (1998). Memory and attention make smart word learning: An alternative account of Akhtar, Carpenter, and Tomasello. *Child Development, 69*, 94-104.

Sanford, A. J., & Garrod, S. C. (1981). *Understanding written language.* Chichester: Wiley.

Sedivy, J. C., Tanenhaus, M. K., Chambers, C. G., and Carlson, G. N. (1999). Achieving incremental semantic interpretation through contextual representation. *Cognition, 71,* 109-147.

Stevenson, R. J., Crawley, R. A., & Kleinman, D. (1994). Thematic roles, focus and the representation of events. *Language and Cognitive Processes, 9,* 473-592.

Stevenson, R., A. Knott, J. Oberlander & S. McDonald. 2000. Interpreting pronouns and connectives: Interactions among focusing, thematic roles, and coherence relations. Language and Cognitive Processes 15. 225–262.

Theeuwes, J. (1991). Exogenous and endogenous control of attention: The effect of visual onsets and offsets. *Perception and Psychophysics, 49*, 83-90.

Theeuwes, J. (1994). Stimulus-driven capture and attentional set: Selective search for color and visual abrupt onsets. *Journal of Experimental Psychology: Human Perception and Performance, 20*, 799-806.

Tily H, Piantadosi S. 2009. Refer efficiently: use less informative expressions for more predictable meanings.Proceedings of the Workshop on the Production of Referring Expressions: Bridging the Gap between Computational and Empirical Approaches to Reference, 2009.

Tomasello, M. (1995). Joint attention as social cognition. In C. Moore & P. J. Dunham (Eds.), *Joint attention, its origins and role in development* (pp.103-130). Hillsdale, NJ: Lawrence Erlbaum.

Tomlin, R. S. (1997). Mapping conceptual representations into linguistic representations: the role of attention in grammar. In J. Nuyts & E. Pederson (Eds.), *Language and Conceptualization*. Cambridge: Cambridge University Press, 162-189.

Trueswell, J.C., Sekerina, I., Hill, N.M. & Logrip, M.L. (1999). [The kindergarten path effect: studying on-line sentence processing in young children.](http://www.psych.upenn.edu/~trueswel/Trueswell_Papers/kpath.pdf) *Cognition*, 73, 89-134.

Van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York: Academic Press.

Van der Meulen, F. F., Meyer, A. S., & Levelt, W. J. M. (2001). Eye movements during the production of nouns and pronouns. Memory & Cognition, 29, 512–521.

Walker, M. A., Joshi, A. K., & Prince, E. F. (1988). *Centering theory in discourse*. Oxford: Oxford University Press.

Yantis, S. (1993). Stimulus-driven attentional capture. *Current Directions in Psychological Science, 2*, 156-161.

Yantis, S., & Jonides, J. (1990). Abrupt visual onsets and selective attention: Voluntary vs. automatic allocation. *Journal of Experimental Psychology: Human Perception and Performance, 16, 121-134.*

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin, 123*, 162-185.

1. The contrast between endogenous and exogenous attention is superficially similar to the contrast between shared and private attention. Yet the two concepts are orthogonal. Mutually attended things are often related to the goals of the task, for example the ball in a game of tennis, or the words in a conversation. But non-task-related, exogenous cues to attention can also be assumed to result in shared attention. For example, a highly salient public stimulus (e.g., a sudden loud noise) can be assumed to have attracted the attention of everyone present, even if it is not task-related. [↑](#footnote-ref-1)
2. For the 25 trials (2.3% of the data) where there were no looks to either character, the N1 proportion was entered as 0. [↑](#footnote-ref-2)
3. Trials on which there were no looks to either target or competitor in the first second were scored as 0 for Trial-initial attention to target. [↑](#footnote-ref-3)