Cognitive Psychology 70 (2014) 58–81

Contents lists available at ScienceDirect

Cognitive Psychology

journal homepage: www.elsevier.com/locate/cogpsych

The road to understanding is paved with the speaker's intentions: Cues to the speaker's attention and intentions affect pronoun comprehension

Rebecca Nappa [a], Jennifer E. Arnold [b,*]

[a] Harvard University, United States
[b] University of North Carolina, Chapel Hill, NC, United States

ARTICLE INFO

Article history:
Accepted 25 December 2013

Keywords:
Pronoun resolution
Language comprehension
Attention
Common ground

ABSTRACT

A series of experiments explore the effects of attention-directing cues on pronoun resolution, contrasting four specific hypotheses about the interpretation of ambiguous pronouns *he* and *she*: (1) it is driven by grammatical rules, (2) it is primarily a function of social processing of the speaker's intention to communicate, (3) it is modulated by the listener's own egocentric attention, and (4) it is primarily a function of learned probabilistic cues. Experiment 1 demonstrates that pronoun interpretation is guided by the well-known N1 (first-mention) bias, which is also modulated by both the speaker's gaze and pointing gestures. Experiment 2 demonstrates that a low-level visual capture cue has no effect on pronoun interpretation, in contrast with the social cue of pointing. Experiment 3 uses a novel intentional cue: the same attention-capture flash as in Experiment 2, but with instructions that the cue is intentionally created by the speaker. This cue does modulate the N1 bias, demonstrating the importance of information about the speaker's intentions to pronoun resolution. Taken in sum, these findings demonstrate that pronoun resolution is a process best categorized as driven by an appreciation of the speaker's communicative intent, which may be subserved by a sensitivity to predictive cues in the environment.

* Corresponding author. Address: CB #3270 Davie Hall, Rm. 337B, Chapel Hill, NC 27599-3270, United States. Fax: +1 919 962 2537.

E-mail address: jarnold@email.unc.edu (J.E. Arnold).

http://dx.doi.org/10.1016/j.cogpsych.2013.12.003
0010-0285/

1. Introduction

Successful communication relies upon listeners understanding the speaker's intended meaning. A significant component of this task is the correct assignment of reference, as the listener must determine which specific items, people, and events the speaker means to discuss. This task is complicated by the fact that all linguistic referring expressions are ultimately ambiguous. An extreme example of referential ambiguity is presented by pronouns, which occur commonly and yet are highly dependent on the context for their interpretation. Although listeners prioritize finding a referent that matches lexical features of the pronoun, e.g. a female for "she" (Arnold, Eisenband, Brown-Schmidt, & Trueswell, 2000), there still may be many entities available as potential referents.

The question we ask here is how the listener solves this problem of referential identification, and specifically, how it is guided by evidence about the speaker's attention and intentions. We focus on the ambiguous personal pronouns *she* and *he*, in discourse contexts that include more than one character matching the gender of the pronoun. We examine the psychological mechanisms by which listeners identify the speaker's intended referent, specifically the effects of social–communicative cues like pointing and gazing.

Gaze and gestures provide a testing ground for distinguishing several potential mechanisms of pronoun resolution. They are dynamic, transitory cues, and thus do not form part of the preceding discourse context in the same way as textual cues, such as syntactic prominence or recency of mention. Yet at the same time, they are systematically related to the speaker's and listener's attention, both of which are also hypothesized to be constrained by the discourse context.

We consider several explanations for pronoun resolution. Note that these explanations are not mutually exclusive: (1) the Grammatical constraints hypothesis, (2) the Speaker-intention hypothesis, (2) the listener's egocentric attention hypothesis, and (4) probabilistic cues to successful understanding. We examine how social–communicative cues (gazing and pointing) are related to both the speaker's and listeners' attention, and provide evidence of the speaker's referential intentions. These effects are considered against the backdrop of well-known discourse context effects.

There is extensive evidence that pronoun resolution is highly constrained by the discourse context (e.g., Arnold, Eisenband et al., 2000; Clark & Sengul, 1979; Gernsbacher, 1989; Gordon, Grosz, & Gilliom, 1993; Sanford & Garrod, 1981). A number of discourse factors predict which characters will be most accessible to subsequent pronoun resolution, including recent mention, parallelism between pronouns and antecedents, and grammatical and thematic roles of antecedents (inter alia, Ariel, 1990; Arnold, 1998; Gernsbacher & Hargreaves, 1988; Givón, 1983; Grosz, Joshi, & Weinstein, 1995; Gundel, Hedberg, & Zacharski, 1993). For example, in the utterance, "Yesterday, Homer ate breakfast with Bart. He had some eggs," most listeners will interpret the pronoun "he" as referring to Homer, since Homer is the first-mentioned character and the subject in the preceding sentence, as well as the referent that is in the parallel syntactic position to the referring pronoun, which is also in subject position.

The most common explanation for these discourse constraints is that some information in the context is salient, and thus is in the focus of attention of all discourse participants (Brennan, 1995; Chafe, 1994; Grosz et al., 1995; Gundel et al., 1993; see Arnold, 2010, for a review). The kind of focus that matters to pronoun interpretation is related to indicators of topicality (e.g., Ariel, 1990; Givón, 1983), as opposed to the linguistic category of focus, which tends to denote the new or focal part of a sentence (Arnold, Kaiser, Kahn, & Kim, 2013). There is ample evidence that discourse focus is a strong determinant of listeners' preferences for pronoun referents, but there are numerous processes that are likely to correlate with discourse focus. Here we outline four mechanisms that could underlie the effects of discourse focus on pronoun interpretation. We then present three experiments designed to tease these mechanisms apart, by examining how listeners' interpretations of pronouns are influenced by the social–communicative cues of pointing and gazing at referents.

1.1. Grammatical constraints

One way of viewing discourse constraints is as a set of rules for how to interpret pronouns and other references. A simplistic rule might be "examine all entities in the preceding clause for a match with the person/number features of the pronoun, starting with the most focused elements". This approach is compatible with many kinds of representations of discourse focus, for example in terms of topicality (Givón, 1983), expectancy (Arnold, 1998), or as a direct compilation of text-based heuristics like recent mention or mention in prominent linguistic positions. The critical feature of the grammatical constraints view is that it draws on a selection process where referent identification is conditioned on rules about the discourse or other criteria. Under this view, the effects of discourse focus are conventional, based on language-specific rules about what is appropriate.

One formalism that takes this approach to pronoun interpretation is based on Centering Theory (Grosz et al., 1995). Although Centering Theory was originally developed to account for what makes a discourse coherent (which is more of a production question), extensions have used it to model pronoun interpretation (e.g., Brennan, Friedman, & Pollard, 1987; Tetrault, 2001; see Kehler, 1997, for a critical review). Centering Theory suggests that discourses are most coherent when pronouns are used to refer to the backward-looking Center (i.e., the discourse focus). The broader goals of Centering Theory, which is based on an earlier model of Grosz and Sidner (1986), are to link discourse structure to representations of the intentional and attentional structures of the discourse. But critically, Centering does so via a computational mechanism that makes predictions based only on the discourse context, and critically the grammatical role in which entities were last mentioned. Entities are assumed to fall along the grammatical hierarchy subject > object > oblique (Brennan et al., 1987), so pronouns are preferentially assigned to the referent of the subject of the previous clause. This view is supported by research on pronoun production, which has found strong effects of the linguistic context on speakers' choices about when to use pronouns (Arnold, 1998, 2008; Fukumura & van Gompel, 2010; Kehler, Kertz, Rohde, & Elman, 2008; Stevenson, Crawley, & Kleinman, 1994). The same discourse cues also strongly constrain comprehension (Gordon et al., 1993; Hudson-D'Zmura & Tanenhaus, 1998).

In sum, the grammatical rules view is one that uses any rule (or set of rules) to explicitly condition the interpretation of a pronoun, for example on the discourse context or other conditioning factors. Under this view, social communicative cues like pointing seem to require a separate kind of rule. This is consistent with a distinction that is frequently made between deictic pronouns and anaphoric pronouns. Although this distinction is most clearly made between pronouns like *this*/*that* and personal pronouns *he*, *she*, *it*, even personal pronouns can be used deictically. This distinction could easily be captured by a rule that states that deictic gestures determine the referent of a concurrent pronoun, and otherwise rules about the discourse context are invoked.

1.2. Joint attention and the speaker's intentions

Another view is that the discourse context is important because it provides one indicator of what the speaker considers to be in joint attention. Many researchers have highlighted mutual perspective taking and the establishment of common ground (i.e., shared information) as the foundation for referential communication (among others, Brennan & Clark, 1996; Clark & Krych, 2004; Gundel et al., 1993; Pickering & Garrod, 2004). Speakers and listeners follow guidelines about how to refer to elements in a discourse, like using definite reference only when referring to objects that are uniquely identifiable.

This view is supported by evidence that pronouns can be used successfully even without the support of an immediate discourse context, as when a friend asks another "So, did he call you?" (Gerrig, Horton, & Stent, 2011; Greene, McKoon, Gerrig, & Ratcliff, 1994). In some cases, deictic cues support the intended interpretation, as when the speaker points and says "What's this?"

A strong cue to the speaker's intentions is likely to be the speaker's attention to task-relevant objects, especially public displays of attention that are likely to direct the listener's focus of attention as well. On this account, language comprehension is a matter of assessing what the speaker means, not

what the words themselves mean. The listener should not be tempted to interpret pronouns as referring to those elements that have attracted their own attention, or those elements that are cued reliably by something other than the speaker. Thus, social cues like pointing and gazing themselves might influence pronoun interpretation, but not extraneous cues that direct the listener's attention at the moment the pronoun is encountered.

Note that there are multiple mechanisms by which shared attention might constrain pronoun interpretation. On one hand, it might constrain the listener's discourse representation, such that shared information is more attended, and more mentally accessible (Ariel, 2001; Arnold, 2010; Arnold & Lao, submitted for publication). On the other hand, it might directly affect pronoun resolution at the moment the pronoun is encountered. In this paper we are primarily concerned with the latter, i.e. questions about the processing that occurs concurrently with the pronoun.

1.3. The listener's egocentric attention

Although communication is clearly about the listener understanding the speaker's intentions, it is also possible that pronoun resolution is guided by the listener's own attention at the moment of encountering a pronoun. This view is a natural consequence of the assumption that pronouns are frequently coreferential with the focused portion of common ground. If discourse focus tends to correlate with pronoun referents, the language comprehension system may have developed an efficient mechanism by which listeners automatically associate pronouns with things in their own focus of attention. Under this view, both the discourse and social cues are relevant because they direct the listener's attention.

It is important to point out that this account is unlikely to be the only explanation, since communication would fail if listeners were to always associate pronouns with any old thing they happened to be thinking about. However, referential communication experiments have demonstrated that listeners do not always track common ground information perfectly (Hanna, Tanenhaus, & Trueswell, 2003; Keysar, Barr, Balin, & Brauner, 2000), especially when under time pressure (Horton & Keysar, 1996). Thus, listeners may use their own attention as a proxy for joint attention. If so, we might expect that the listener's egocentric attention could have a partial effect on pronoun resolution, even if it is not the primary constraint.

As with shared attention, egocentric attention could affect pronoun comprehension in several ways. One is by modulating the accessibility of information in the listener's mental representation, and indeed there is evidence that it does (Arnold & Lao, submitted for publication). By contrast, our question in this paper is whether pronoun comprehension is guided by the listener's attention at the moment the pronoun is encountered. This question is especially relevant to social cues like pointing, since they may guide the listener's attention, in addition to any other information they carry.

1.4. Probabilistic cues to successful understanding

A fourth view is that both discourse effects and the effects of social cues on pronoun resolution are largely learned contingencies between contextual cues that speakers tend to provide, and the elements referential terms ultimately refer to (Arnold, Brown-Schmidt, & Trueswell, 2007). Speakers do frequently continue to talk about items that have been focused in the discourse, and often point to objects they wish to refer to. On this account, the relationship between a cued item (e.g. a dog that's been pointed to) and a spoken reference (e.g. "that dog") is an association that listeners learn to make, simply because it reliably predicts how words will refer. No mental modeling of the speaker's intentions is required here, as the relationship between objects and contextual cues to reference is simply a heuristic, grown out of the way speakers have tended to use these cues in the past. Under this view, listeners can learn the predictive power of any contextual cue, including discourse context and speakers' actions. However, the listener's own focus of attention is unlikely to matter, as the listener's attention is unrelated to the probabilistic relationship between a contextual cue and a potential referent (assuming enough attention is dedicated to processing the language that the cue is not missed entirely).

2. The time course of attentional and intentional cues to pronoun resolution

Our study tests the above explanations by examining cues like pointing and gazing. Critically, we are interested in the effect of these cues have at the moment when a pronoun is encountered. That is, we ask whether signaling the appropriate referent (e.g., by pointing) at the moment of uttering a pronoun influences pronoun resolution processes.

We distinguish these from processes of discourse comprehension that occur before the pronoun is encountered, during which listeners use the structure of a sentence to establish a mental representation of the discourse, in which some entities are more accessible than others. Thus, as the listener hears *Bunny is playing with Froggy*, she instantiates a mental representation of the characters Bunny and Froggy, and the playing event that connects them. Given the accessibility of the grammatical subject position, Bunny is likely to receive a representation that is more accessible than Froggy.

Several theories propose that accessibility itself (as opposed to pronoun resolution) is related to attention (e.g. Brennan, 1995; Gundel et al., 1993). As Brennan (1995) proposes, when a speaker mentions an entity as the grammatical subject, it sends a message to the listener "my attention is here." The speaker elects to highlight certain discourse entities (by mentioning them often or placing them in more prominent linguistic positions, for example), thus directing more attention to these items and making them more accessible to both speaker and listener for subsequent processing. Under this view, accessibility is conceptualized as increased attention to focused characters, and as a result, the mechanism for resolving reference may rely heavily on the listener's attentional state.

Empirical findings support the hypothesis that discourse accessibility is related to attention. For example, Gleitman, January, Nappa, and Trueswell (2007) showed that the speaker's attention to a discourse scene affects the speaker's choices about how to frame a description of a scene. On the comprehension side, Arnold and Lao (submitted for publication) found that when listeners allocate visual attention to a pictured character early in a story, they later show a preference for considering that character during pronoun resolution. Critically, both the production and comprehension effects were linked to egocentric attention, rather than any public signal about discourse accessibility.

These lines of research have largely demonstrated that the activation of conceptual information via language or visual input leads to increased accessibility in the other domain as well – i.e. visually attending to an entity makes it easier to retrieve linguistically, and items that are activated linguistically tend to attract visual attention (see also Salverda & Altmann, 2011). But importantly, this work has centered around the discourse construction stage: as one builds a visual representation of events, and/or a linguistic representation of a discourse, low-level aspects of one's own attention influence the process. Relatively little work has explored attentional factors at the reference resolution stage (at the point when the listener encounters an ambiguous referential term).

However, a few studies do demonstrate effects of social communicative cues, like pointing and gazing, which may influence both the listener's attention and their assessment of the speaker's intentions. For example, Hanna and Brennan (2008) found that the speed of reference comprehension was influenced by the speaker's eyegaze during the production of temporarily ambiguous noun phrases like *the blue circle with five dots*. In addition, Goodrich and Hudson Kam (2009, 2012) found that pointing gestures affected pronoun interpretation. In their experiments, participants watched videos of a speaker who used gestures as they constructed the discourse situation, for example indicating one character on the left side, and one on the right. In this context, a gesture at the start of the pronoun influenced listeners' interpretation of the pronoun. Gesture has also been studied as it relates to joint attention and communication in young children (Baldwin, 1995; Carpenter, Nagell, & Tomasello, 1998). However, none of these studies examined how gaze and gesture interacted with discourse cues. In addition, these findings are consistent with both the hypothesis that listeners seek out cues to the speaker's intentions, and the hypothesis that these cues guide the listener's own egocentric attention.

3. The current study

The following experiments compare different explanations of how listeners interpret pronouns in a spoken discourse. We use the same task as Arnold et al. (2007, Experiment 1), in which listeners hear a very short story about two toy characters, e.g. *Bunny is playing with Froggy. She wants the ball.* The listener's job is to say who wants the ball, Bunny or Froggy. This task was originally developed for use with children, but here is adapted for experiments with adults. The first sentence mentions two characters, which creates a situation in which the first-mentioned character (N1) is deemed more focused in the discourse, and more accessible to adult listeners (Gernsbacher & Hargreaves, 1988; Gordon et al., 1993). This focus then influences the interpretation of subsequent pronouns. This N1 bias is particularly evident in a context where the two characters are the same gender, where the critical pronoun is ambiguous. For example, adults in the Arnold et al. (2007) task tended to prefer the first-mentioned character (Bunny) as the referent, choosing it 88% of the time, but not as categorically as they did when only the first character matched the gender of the pronoun, in which case they chose it 100% of the time.

This paradigm provides an excellent testing ground for exploring the effects of social communicative cues against the backdrop of the reliable, yet probabilistic discourse constraints from order of mention. Experiment 1 examines how the N1 bias is modulated by two social cues that occur at the moment the pronoun is uttered: eye gaze, as a cue to the speaker's focus of attention, and pointing, as a cue to directed attention. Experiment 2 tests the hypothesis that social cues affect pronoun resolution as a result of orienting the listener's egocentric attention automatically to one object. Experiment 3 tests the hypothesis that listeners track information about the speaker's referential intentions, as opposed to an automatic use of learned contingencies.

4. Experiment 1

Experiment 1 examined listeners' preferences during the interpretation of ambiguous pronouns, contrasting first-mention bias effects with both (1) the effects of the speaker's attention (manipulated by way of gaze cues) and (2) directed attention (manipulated by way of combined pointing/gaze) on reference resolution. This experiment tests the relative strength of each type of cue, against the well-known first-mention bias effect.

The gaze condition is particularly important for testing the hypothesis that pronoun resolution is primarily driven by grammatical constraints. It is plausible that the effects of both the first mention bias and pointing gestures are governed by rules about how to interpret pronouns. However, it is less plausible that rules condition pronoun interpretation choices on the basis of gaze cues per se. If the speaker's gaze affects listener's understanding of the speaker's meaning, it would be a strong indicator that pronoun resolution goes beyond discourse rules.

4.1. Methods

4.1.1. Participants

A total of 47 native speakers of English participated at the University of North Carolina, Chapel Hill, in exchange for course credit. Five were excluded due to technical problems with the experimental program. This left 42 participants (7 in each of the 6 lists) in the final analysis.

4.1.2. Task overview

The same task and stimulus structure was used as in Arnold et al. (2007), except that all stories were pre-recorded on a video rather than performed live. Each participant viewed a total of 58 videos (24 experimental stimuli, 34 fillers). In each video (see Fig. 1), participants saw a woman (the first author) sitting at a table with two toy animal characters, one on either side of her, and a toy object in the center of the table. The woman in the video introduced the two toy characters and told a simple story about them, and asked a question about the object, e.g.:

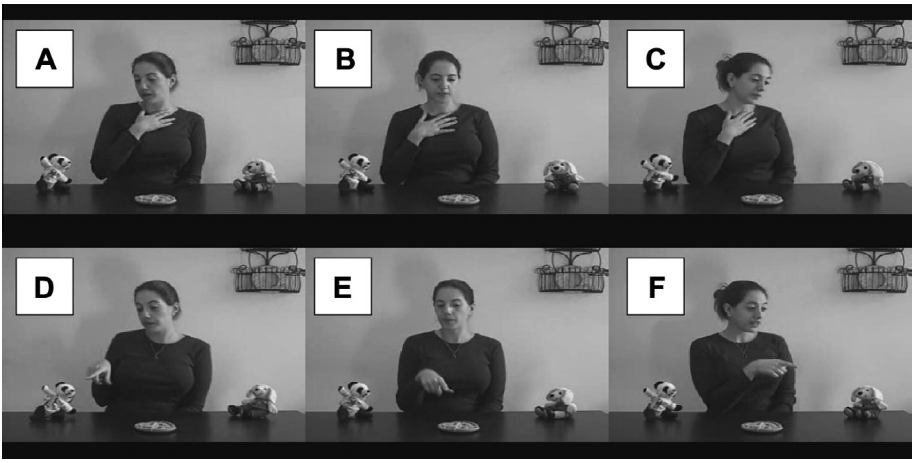

Fig. 1. Photographs of conditions 1 through 6 in Experiment 1. Panels (A) through (C) depict the Gazing-only conditions: A = gazing at N1, B = neutral gaze, C = gazing at N2. Panels (D) through (F) depict the Pointing-and-Gazing conditions: D = pointing to N1, E = neutral pointing, F = pointing to N2.

This story is about Puppy and Panda Bear. Puppy is having some pizza with Panda Bear. He wants the pepperoni slice.

The participant's job was to indicate which character wanted the mentioned object (e.g., the pepperoni slice) by pressing a button. We analyzed both (1) the proportion of N2 responses, i.e. those that go against the discourse-based N1 bias, and (2) the latency to respond in each condition.

4.1.3. Stimuli

One hundred and forty-four videos were created, with each of 24 stories in each of six conditions (see Fig. 1 for an example). Sample videos are available at http://arnoldlab.web.unc.edu/publications/supporting-materials/nappa-arnold/. See the Appendix for all stimuli and fillers.

Condition A – Looking cue to N1 (first-mentioned character in preceding sentence).
Condition B – Neutral looking cue (to a central location).
Condition C – Looking cue to N2 (second-mentioned character in preceding sentence).
Condition D – Pointing and Looking cue to N1 (first-mentioned character in preceding sentence).
Condition E – Neutral pointing and looking cue (to a central location).
Condition F – Pointing and Looking cue to N2 (second-mentioned character in preceding sentence).

Filler videos followed a similar format as the experimental stimuli, except the two characters sometimes were of different genders, and the second sentence usually mentioned one of them by name.

Participants saw each story once, in one of the 6 conditions, with four stories per condition, creating a 2×3 factorial design with Cue Type (Pointing-and-Looking, Looking-Only) crossed with Cue Location (N1, N2, or a central Neutral location). The timing of the critical gestures and gazes overlapped with the production of the pronoun. Note that all pointing gestures were accompanied by gazes to the same location to preserve naturalness, so the critical contrast is between Looking-Only and Pointing-and-Looking. Participants also saw 34 filler videos, for a total of 58 videos viewed by each subject. In each video, two characters were introduced. In all target videos, the two characters were the same gender for all target videos, and the second sentence began with an ambiguous pronoun (*He* or *She*).

Location of N1 was counterbalanced, such that N1 appeared on the left on half of the 4 trials in each condition, and on the right on the other half, to control for any left- or right-hand side of the screen biases. The locations of the characters onscreen were controlled across videos, such that characters always subtended an equivalent visual angle, and were located in the same area of space.

Each video was approximately equivalent in length, ranging from 12 s to 15 s, with slight variations due to the length of the predicate (e.g. *Puppy is getting ready for school with Panda Bear. Panda Bear wants to wear the blue shoes,* would be slightly longer than the story about Bunny wanting the ball above.) All videos were recorded using a Sony digital video camera, and subsequently edited in Windows Movie Maker.

4.1.4. Procedure

At the outset of the experiment, the experimenter introduced and indicated the gender of all 4 characters (Puppy and Panda Bear, who were identified as male, and Bunny and Froggy, who were identified as female). Participants were told that a subset of these videos could subsequently be used with preschool-aged children, to establish the plausibility of using childlike stories. Participants were then told that they would view a series of stories on video, which would each be followed by a memory question, to ensure that they were attending to the stories in the videos, and a plausibility question, to assess how reasonable they thought it was that these characters would engage in the action described. For target trials, the "memory question" assessed pronoun resolution. For example, for the item displayed in Fig. 1, the question was, *Who wants a pepperoni slice?* For filler items, the memory question was either about location (e.g., *Who was on the left hand side of the screen?*) or a question like *What does Froggy want to eat?* For each item, the memory question was then followed by a plausibility rating, for the purpose of drawing participants' attention away from the purpose of the experiment. These ratings were not analyzed.

The experiment was presented using a software package called Behavior and Cognitive Research (BCR), developed at the University of North Carolina at Chapel Hill (Pitz, 2007). Participants viewed the videos on 17-inch flat panel monitors, seated approximately 18 inches from the screen. Each trial included (a) a video, (b) a screen with a memory question, and (c) a screen with a plausibility question. Participants pressed a button to move on from one trial to the next, and then to select the character they believed to be the correct answer, and to assess plausibility. Responses to the memory question and latencies to make this response were recorded and analyzed.

4.1.5. Data analysis

We analyzed two primary dependent measures: (a) the response to the memory question as a measure of their interpretation of the pronoun, and (b) their latency to respond, measured from the onset of the question screen. We analyzed responses with a multilevel logistic regression, using SAS proc glimmix with a binary distribution and a logit link. We analyzed latencies with a multilevel mixed effects model, using SAS proc mixed. For the latency analysis, outliers that were more than 3 times the standard deviation from the mean were excluded ($n = 6$ for Exp. 1). For each model, we first built a model with only the control variables; any that contributed at $t \geqslant 1.5$ were retained for the final model.

All models included crossed random effects for participant and item. If not otherwise mentioned, all models also included random slopes for critical predictors with respect to both participant and item. In some cases, the model returned an error that the G matrix was not positive definite; in this case we removed each random slope (starting with those for the items) until we found a model that did converge. The random effects structure is noted for any model on which the structure was less than full.

4.2. Results

The participants' responses revealed that, as expected, participants were much more likely to arrive at an N1 than an N2 interpretation in the baseline Neutral conditions (~20% N2 interpretations; see Fig. 2). These preferences were modulated by both pointing and gaze cues. When the speaker pointed or looked at N1, this preference for an N1 interpretation increased (only ~5% N2 interpretations).

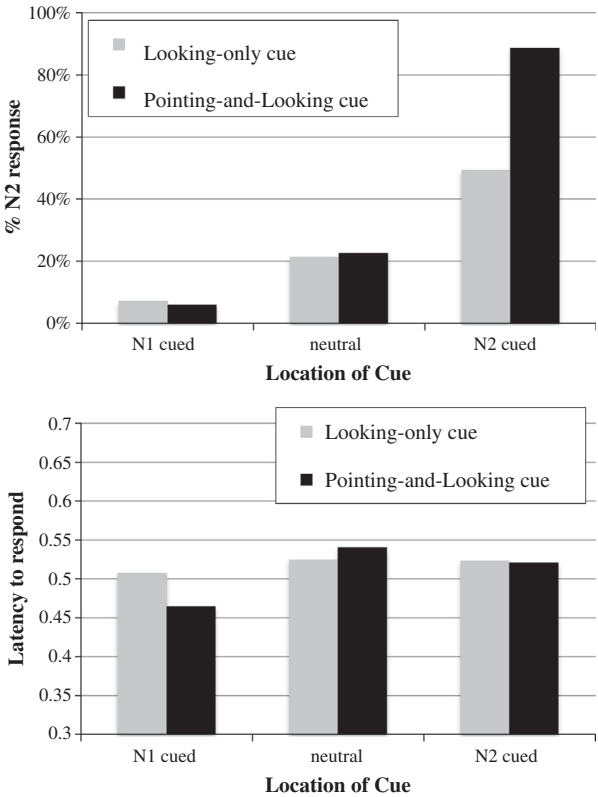

Fig. 2. Results of Experiment 1: proportion of N2 responses (top panel) and Latencies to respond to the pronoun interpretation question (bottom panel) when pointing and gaze cues were to N1 (first-mentioned character in preceding sentence), the object (neutral), or N2 (second-mentioned character in preceding sentence).

When the speaker pointed at N2, participants reversed this N1 bias, and chose N2 on a majority of trials (~90%). Gazing at N2 also disrupted the N1 bias, but not as strongly as the pointing cue, leading to an N2 interpretation on ~50% of the trials.

These findings were supported by the mixed-effects logistic regression (see Table 1). There was a marginal main effect of Cue type, a main effect of Cue location, and a significant interaction between the two.

Cue location was a three-way manipulation. Planned comparisons among these conditions were conducted using Simple Contrast coding. In one model, we compared N1 to the neutral condition,[1] and found a main effect of Cue type ($\beta = .62$ (0.14); $t = 4.46$, $p < .0001$), and a main effect of cue location ($\beta = -0.38$ (0.13); $t = -2.98$, $p = .0029$), but no interaction. In another model, we compared the neutral to the N2 condition, and found main effects of both Cue type ($\beta = 0.54$ (SE = 0.17); $t = 3.22$, $p = 0.002$), Cue location ($\beta = -0.87$ (SE = 0.22); $t = -3.94$, $p = 0.0004$), and an interaction between the two ($\beta = -1.05$ (SE = 0.22); $t = -4.83$, $p < .0001$). This means that cues to N1 of either type boosted the selection of that character, compared to the neutral condition, and they did so equally. Cues to N2 also differed from the neutral condition, but not to the same degree. This reflects the fact that the pointing-and-gazing at N2 cue was stronger than the gazing-only at N2 cue.

[1] In the first model, the cue-to-N1 condition was coded as 1, and compared to the neutral condition, which was coded as −1. The N2 condition was coded as 0. This model had only random intercepts, no random slopes. In the second model, the cue-to-N2 condition was coded as -1, the neutral condition as 1, and the N1 condition as 0. This model had random intercepts and random slopes for the cue location (by both participants and items), but no random slopes by cue type.

Table 1
Statistical parameters and Effect sizes for Experiment 1 analyses. The dependent variable for responses is N1 selection (vs. N2), and for latencies is the log of the latency between the onset of the question screen and the time of response.[a]

Effect	Responses			Latencies		
	β (Error)	t	p	β (Error)	t	p
Cue location (N1 vs. neutral vs. N2)	−1.41 (0.21)	−6.55	<.0001	−0.01 (0.01)	−0.83	0.41
Cue type (Point-and-Look vs. Gaze-only	0.44 (0.2)	2.23	0.0263	−0.01 (0.01)	−0.99	0.33
Type × Location	−1.53 (0.28)	−5.49	<.0001	−0.02 (0.01)	−1.75	0.08

[a] The responses analysis included random intercepts for both participants and items, and random slopes for cue location by both participants and items, but no random slopes for cue type.

Latency analyses revealed that participants were also faster to respond when the speaker pointed to N1 than in any other condition (see Fig. 2). This effect did not emerge strongly in the full analysis (see Table 1), which revealed only a marginal interaction between cue type and cue location, and no main effects. Analyses with simple contrast coding revealed that the interaction was robust for the N1 vs. control model ($\beta = -0.03$ (SE = 0.01); $t = -2.62$, $p < 0.001$) but not the control vs. N2 model ($\beta = 0.01$ (SE = 0.01); $t = 0.81$, $p = 0.42$); none of the main effects were significant in either model.

4.3. Discussion

These results show the substantial effects of attention-directing cues on the resolution of referential expressions, particularly for the combined Pointing-and-Looking gesture. Both deictic gesture cues and gaze cues to N1 enhanced the effects of the prior discourse (order-of-mention), bolstering participants' baseline tendency to interpret the ambiguous pronoun as referring to N1. Cues to N2 significantly increased participants' tendency to interpret the ambiguous pronoun as referring to N2. The effects of pointing-and-looking were stronger than those of the looking-only cue.

Clearly, listeners factor both gaze and gesture cues into referential assignment decisions, but only explicit information (from gesture cues) trumped the reliable order-of-mention cue in the linguistic input. Response times largely agreed with these findings – responses were speeded when gesture cues are to N1, but otherwise showed no effect of attention-directing cues. Gaze cues had little effect on response times. Even gaze cues that agreed with the order-of-mention information did not expedite the pronoun resolution process (despite increasing the likelihood that participants would arrive at an N1 interpretation). And interestingly, gesture or gaze cues that conflicted with prior discourse information (order-of-mention) did not slow participants down significantly as they make their referential responses.

These results add to a growing set of findings that pronoun resolution is not driven by the linguistic context alone. While the discourse context had a robust effect, it was modulated by eye gaze cues about the speaker's attention, and even more so by explicit cues about where the listener is meant to direct their attention (i.e. the element the speaker intends to refer to). Even if that element is not the one the listener would otherwise have expected the speaker to reference, this conflict was quickly resolved in overwhelming favor of the cue to the speaker's intentions.

These findings do not fit easily with an account of pronoun interpretation that is based on rules. It is plausible that our linguistic competence includes rules for pronouns that are conditioned on the discourse context, e.g., based on discourse focus. Under this view, deictic gestures like pointing may engage additional rules that override the discourse. Yet gaze does not lend itself to a grammatical approach. Speakers look at many things as they speak, and not always the things they refer to. As our data revealed, gaze alone is a relatively weak cue, and does not overturn the discourse context. If anything, gaze may serve as a probabilistic constraint.

Thus, our findings fit well with assumptions that discourse constraints are relevant because they tap into the speaker's intentions and joint attention (e.g., Chafe, 1994; Clark & Marshall, 1981; Grosz & Sidner, 1986), and not just as grammatical constraints. This finding is consistent with the idea that discourse focus is not merely a linguistic category, but is in fact related to calculations about the speaker's attention and intentions (Arnold & Lao, submitted for publication; Brennan, 1995).

5. Experiment 2

Experiment 1 suggested that listeners are sensitive to social cues about the speaker's attention and intended reference, and use these to interpret pronouns. Yet there are multiple psychological mechanisms that could underlie these effects. Notably, pointing and gazing cues could affect the listener in two general ways.

On the one hand, listeners could solve the pronoun resolution task by attempting to determine the communicative goals of the speaker, factoring in information like common ground, keeping careful track of what the speaker knows, and assessing what she intends to communicate as she speaks. This could be what speeds the pronoun resolution process and definitively pushes pronoun interpretations in whichever direction the speaker indicated (either N1 or N2) when the speaker provides explicit gesture cues to her intentions. Under this view, pointing is a conventional gesture that indicates the deictic interpretation of the pronoun. By contrast, gaze cues (without pointing) allow the listener to construct a representation of the speaker's attention, and thus infer the intended meaning. This analysis of Exp. 1 results could also explain why the implicit cue of gaze direction, which simply indicates the location of the speaker's attention as she utters the pronoun, does not speed interpretations of any kind, and is often not strong enough to override order-of-mention biases. The listener may simply not consider gaze direction alone to be a reliable enough cue to intention – after all, speakers do not always look at what they're referring to (and vice versa), so this cue is probabilistic at best. In sum, all the results from Experiment 1 could be explained by appealing to a model of language comprehension that is driven by a speaker's goal to communicate.

On the other hand, it is plausible to imagine that the stronger effects of the Pointing-and-Looking cue do not result from any sort of intention-reading success on the part of the listener, but rather stem from the fact that the added physical gesture was simply a more effective and efficient way of directing the listener's *attention* to the item in question. Perhaps the listener is not assessing the speaker's communicative goals at all, but rather is using a simpler heuristic of attention to resolve ambiguous pronouns. Increased attention to a given character could lead to increased activation of the listener's representation of that character, producing an increased likelihood of interpreting the pronoun accordingly. This kind of mechanism would follow naturally from evidence that low-level attention and simple salience affect ambiguity resolution in visual processing (Georgiades & Harris, 1997), and that pointing cues lead to automatic orienting (Langton & Bruce, 2000). It would also be consistent with evidence that the listener's focus of attention influences word ordering in production (Gleitman et al., 2007), and the perceived salience of discourse characters during pronoun resolution (Arnold & Lao, submitted for publication). Thus, perhaps listeners interpret pronouns based simply (or in part) on where their attention is allocated as they interpret the pronoun.

Experiment 2 explores the relative roles of egocentric attention and speaker-centered intention-reading in a similar pronoun resolution task, to attempt to tease these two explanations apart. To assess the contribution of egocentric attention, we adopted a manipulation of automatic visual attention capture (see Georgiades & Harris, 1997 for a discussion of attention-capture, or Gleitman et al., 2007 for an example of how this method has been used in psycholinguistic research). Such cues generally enhance attention in the spatial location where they have occurred (e.g. past studies have found that participants will foveate the cued location 75–80% of the time), but are often subliminal, and should in no way be interpreted as related to goals of the accompanying utterance or the attentional state of the individual producing it. Thus, a comparison of this cue with the pointing cue will allow us to see if effects of the pointing cue on pronoun resolution are the result of some processing of the speaker's communicative goals, or are simply occurring by way of enhanced attention to the cued character. If we see an effect of low-level visual attention on pronoun resolution as well, this will be indicative of a role for general attention in the activation of potential referents as a pronoun is resolved. If such an effect is not found, this will indicate that referential ambiguity is resolved based on the listener's assessments of the elements the speaker *intends* to highlight.

To ensure that attention was, in fact, directed to the attention-capture cue, and as an additional point of comparison between these two types of cue, we asked two different questions following the video. Half of all items were followed by a pronoun resolution question (e.g. *Who wants to play with the ball*), and the other half were followed by a general, visuo-spatial attention-measuring

question: *Who was on the left-/right-hand side of the screen?* The visual-spatial question was aimed at assessing levels of attention to the cued character, as a means of determining whether the attention-capture cue did, in fact, enhance attention to the cued entity.

This question additionally served as a point of comparison between the two types of cues, since it assesses *only* the level of attention allocated to the character in question. That is, the question is whether both pointing and attention-capture cues enhance attention to the cued item equally. If so, the question *Who was on the left side?* should be equally facilitated by both pointing at the left character, and an attention capture cue on the left side. If pointing and attention-capture manipulations lead to different rates of accuracy on this question, it would indicate differing attentional mechanisms for explicit, attention-directing cues (pointing) and low-level, visual attention-capturing cues.

Thus, Experiment 2 allows us to compare the relative contributions of directed attention (pointing gestures) and low-level visual attention (attention-capture) to both pronoun resolution and responses to visuo-spatial questions about a display.

5.1. Methods

5.1.1. Participants

A total of 26 native speakers of English participated at the University of North Carolina, Chapel Hill, in exchange for course credit. Two were excluded due to issues with the experimental program. This left 24 participants (6 in each of the 4 lists) in the final analysis.

5.1.2. Stimuli

The stimuli depicting Condition D and Condition F (pointing cues to N1 and N2, respectively) from Experiment 1 were used in this experiment, along with stimuli in which an attention-capture manipulation drew participants' visual attention to either N1 or N2 (see Fig. 3). These attention-capture flash stimuli were created by adding a small black square to the videos that had been in Condition E (a neutral pointing cue), which appeared as the pronoun was uttered, and was onscreen for approximately 50 ms. Thus, in Experiment 2, there were four conditions:

Condition D – Pointing and Gazing cue to N1.
Condition F – Pointing and Gazing cue to N2.
Condition G – Attention-capture flash cue to N1 (Point and Gaze directed at neutral object).
Condition H – Attention-capture flash cue to N2 (Point and Gaze directed at neutral object).

Each subject again saw 58 videos in total, the same 34 fillers used in Experiment 1, and 24 target videos, 6 in each of the above conditions, with identity (e.g. Puppy vs. Panda Bear) and location of N1 (on the right- vs. the left-hand side of the screen) counterbalanced within each condition.

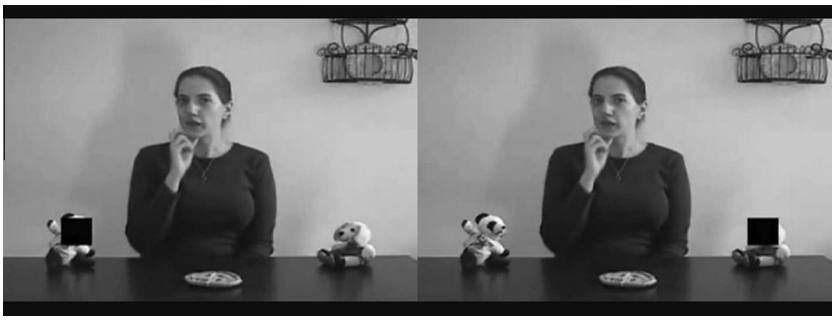

Fig. 3. Photographs of conditions G (left) and H (right) in Experiment 2, with the black attention-capture flash over N1 and N2, respectively.

A third manipulation was the type of question asked. For 12 items, we asked a visuo-spatial attention question (*Which character was on the left-/right-hand side of the screen?*); and for the other 12, we asked the pronoun question. The division of items by question type resulted in 8 conditions: for each set of 12 items, Cue Type (Pointing vs. Flash) and Cue Location (N1 vs. N2) were manipulated within items. Question Type (Visuo-Spatial vs. Pronoun Resolution) was manipulated between items. Across these conditions, identity and location of N1 was split evenly, but these factors were not fully counterbalanced within each condition, as there were only 3 items in each condition. So, an equal number of items had each character as N1, and had N1 on the left, but within any given condition, there may have been only 3 characters represented as N1, and two videos may have had N1 on the left, for example. Items were rotated through each of these conditions and combined pseudorandomly in 8 lists.

5.1.3. Procedure

The procedure was identical to that of Experiment 1.

5.1.4. Data analysis

Data were first split into two subsets, as a function of which Question Type was asked; it would be difficult to consider what a main effect of Question Type would mean in an analysis of all items together, since one question type has a distinct correct answer and the other is a referential interpretation. The same analytical approach was used as for Experiment 1.

5.2. Results

5.2.1. Pronoun resolution

As shown in Fig. 4, results from the pointing gesture conditions in Experiment 1 were replicated: participants' interpretations of the pronoun rarely disagreed with the order-of-mention bias when pointing cues were to N1 (~5% N2 interpretations), but this order-of-mention bias was again reversed when pointing cues were to N2 (~85% N2 interpretations). Results from the attention-capture manipulation, however, showed no such effects: rates of N2 interpretations were around baseline, regardless of the location of the cue (~15% N2 interpretations). These findings were reflected in the statistical analysis (see Table 2), which revealed main effects of cue location and cue type, and an interaction between the two. Planned comparisons[2] demonstrated that Pointing-and-Looking cues yielded a main effect of Cue location ($\beta = -5.31$ (SE = 0.73); $t = -7.26$, $p < .0001$), but the Attention flash cues did not ($\beta = -0.23$ (SE = 0.48); $t = -0.49$, p 0.6294).

Latency analyses showed a main effect of cue location, demonstrating that responses were faster following cues to N1 than to N2. Although this did not interact with cue type, planned comparisons revealed that the effect was stronger for the Pointing cues, for which there was a significant effect of Cue location ($\beta = -0.1$ (0.03); $t = -3.49$, $p = 0.008$), while for the attention-capture cues the effect failed to reach significance ($\beta = -0.05$ (0.03); $t = -1.73$, $p = 0.12$).

5.2.2. Visuo-spatial attention

Results from the visuo-spatial attention question (e.g. "Which character was on the lefthand side of the screen?") demonstrated that participants were highly accurate across the board (see Fig. 5). However, correct answers were slightly more likely when the low-level visual cue indicated the correct character. This was not the case for the pointing cue, which yielded slightly fewer correct answers when the correct character was cued. This emerged as an interaction between Cue location and Cue type, but no main effects of either (see Table 3). However, the effect of correct vs. incorrect cue location failed to reach significance in planned comparisons for both pointing cues ($\beta = 1.0$ (SE = 0.61); $t = 1.65$, $p = 0.10$) and attention-capture flash cues ($\beta = -1.1$ (SE = 0.71); $t = -1.56$, $p = 0.13$). There were no effects on the latency to respond.

[2] The Pointing-and-Looking analysis included only a random intercept for participants, and the Attention flash analysis included random intercepts for both participants and items, and a random slope for Cue location by participants but not items.

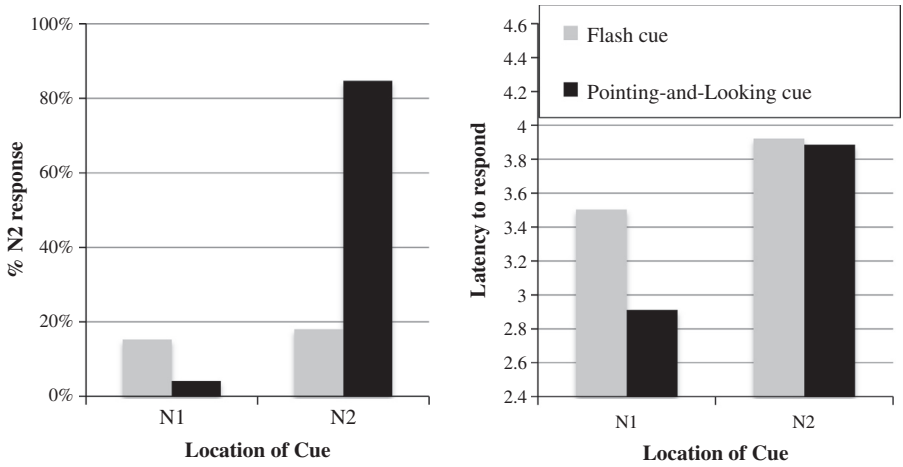

Fig. 4. Results of Experiment 2: proportion of N2 responses (left panel) and latencies (right panel) when pointing and attention-capture cues were to N1 (first-mentioned character in preceding sentence) and N2 (second-mentioned character in preceding sentence).

Table 2
Statistical parameters and Effect sizes for Experiment 2 analyses of pronoun interpretation items.[a]

Effect	Responses			Latencies		
	β (Error)	t	p	β (Error)	t	p
Cue location (N1 vs. neutral vs. N2)	−6.3 (0.83)	−7.64	<.0001	−0.1 (0.03)	−3.57	0.0072
Cue type (Point-and-Look vs. Flash)	−4.26 (0.57)	−7.41	<.0001	0.01 (0.03)	0.38	0.7123
Type × Location	5.89 (0.93)	6.31	<.0001	0.05 (0.04)	1.3	0.1959
Cue on right vs. left	−0.43 (0.44)	−0.98	0.327	NA	NA	NA

[a] The responses model contained only random intercepts for participants and items.

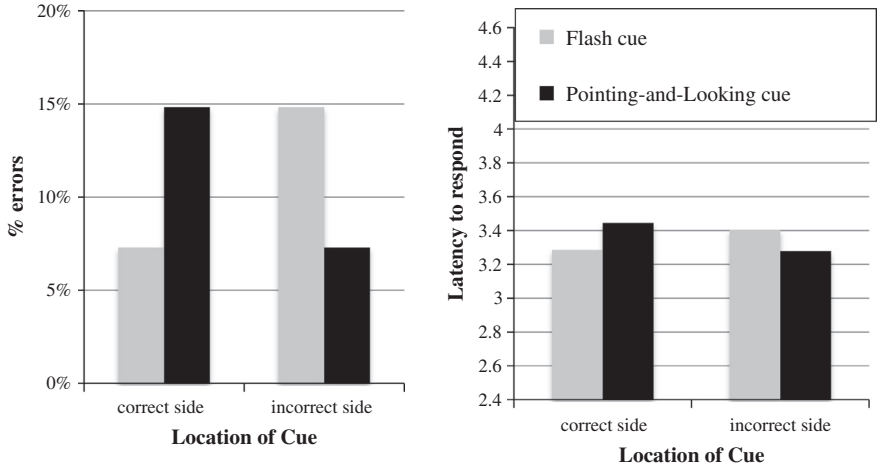

Fig. 5. Proportion of incorrect responses to the visuo-spatial question (e.g. "Which character was on the lefthand side of the screen?"; left panel) and latencies to respond (right panel) when cues were over the correct character (left character for the left question; right character for the right question) and the incorrect character (e.g., right character for the left question).

Table 3
Statistical parameters and Effect sizes for Experiment 2 analyses of visuo-spatial question items.

Effect	Responses[a]			Latencies		
	β (Error)	t	p	β (Error)	t	p
Cue Location (to correct vs. incorrect answer)	1 (0.61)	1.63	0.1033	0.02 (0.02)	0.66	0.525
Cue type (Point-and-Look vs. Flash)	0.86 (0.61)	1.4	0.1618	0.01 (0.03)	0.42	0.6819
Cue side × Cue type	−2.11 (0.9)	−2.35	0.0197	−0.02 (0.03)	−0.64	0.5228

[a] The responses model contained only random intercepts for participants and items.

5.3. Discussion

Findings in Experiment 2 replicated findings from Experiment 1: pointing-and-looking gesture cues again had a strong effect on final pronoun resolution, both in terms of final response and speed of responding. Critically, results also demonstrated that pronoun interpretation was completely unaffected by the listener's own egocentric focus of attention at the moment that a pronoun was uttered.

Results of responses to the visuo-spatial question allow us to draw some conclusions about why these gesture cues enhanced the accessibility of a cued character, whereas attention-capture manipulations did not. The attention-capture flash did succeed in somewhat increasing the proportion of correct responses to the visuo-spatial question. This seems to indicate that the attention-capture cue did, in fact, enhance attention to the cued character. However, since attention-capture had no effect on pronoun resolution, it would seem that a simple increase in allocation of visual attention to a character was not sufficient to affect pronoun interpretation.

Given these results, it is reasonable to conclude that referential interpretations are driven by an attempt, on the part of the listener, to infer the speaker's communicative intentions as speech is produced. These referential interpretations cannot be reduced to the listener's own low-level attentional focus, even though social–communicative cues like pointing and gazing also likely direct the listener's attention. Rather, listeners assign pronoun interpretation based on cues that indicate the *speaker's* attentional state (i.e., gaze effects in Exp. 1), and even more by the speaker's *intention* to direct the listener's attentional state through pointing. Without an intentional component, listeners ignore the location of their current attentional state in referential processing.

6. Experiment 3

Experiments 1 and 2 established that both pointing and gazing cues modulated pronoun interpretation. The gaze cues in particular were inconsistent with a rule-based account of pronoun resolution. Experiment 2 ruled out the hypothesis that the listener's attention guided pronoun interpretation. However, these findings are still consistent with both the Speaker-Intention and Probabilistic Cues hypotheses. Experiment 3 is designed to distinguish between these.

Under the Speaker-intention hypothesis, the distinction between pointing, gaze, and attention-capture effects resulted from their utility in signaling the speaker's intentions. The attention-capture cue was unrelated to the speaker's meaning, and thus had no effect. The stronger effect of pointing than gazing can be explained by the fact that pointing introduces an extra level of social cognition, in which listeners gain access to information about where the speaker intends to refer, rather than just information about where the speaker herself was attending. It is possible that having this additional level of information about the speaker's intentions qualitatively differs from just having information about the speaker's locus of attention. That is, explicit intentional information may enjoy special status with regard to referential assignment (given the high relevance of information about the speaker's intentions to the task at hand), and may show such strong effects on pronoun resolution because this level of representation is intimately tied to our concept of what it means to refer.

However, these results are also consistent with the Probabilistic Cues hypothesis: listeners learn to make use of all sources of information that are reliable indicators of the intended referent. Critically, the attention-capture cue in Experiment 2 was not the kind of cue that listeners would have

experience with, and thus could not have learned to use it as an indicator of pronoun interpretation. Thus, perhaps listeners are not specifically attuned to information about the speaker's intentions, but rather are better able to make use of familiar cues.

Moreover, the Probabilistic Cues hypothesis is also consistent with the contrast between gaze and pointing cues in Experiment 2. Notably, the pointing manipulation in Experiment 1 was always accompanied by a gazing cue. This confounding reflects our goal of using naturalistic stimuli: one seldom points at something without looking at it in day-to-day conversation, in that a speaker's intention to refer and her attentional state are bound to be interrelated. This confounding, however, makes it difficult to tell if the effects of pointing are driven by a powerful but independent effect of explicit intentional information on reference resolution, or simply by way of two separate cues coming together to activate the cued potential referent. In other words, it is unclear whether the same effects would have been seen in the absence of the gaze cue, simply by virtue of a cue to the speaker's intended referent, or if they are the result of gaze cues and pointing cues compounding and together overriding prior discourse information.

To distinguish these two interpretations, we developed a novel paradigm for offering participants intentional information about where the speaker means to refer, but without using established conventional mechanisms like gaze or pointing. This was achieved by way of an introductory video instructing listeners to treat the attention-capture flash from Experiment 2 as intentional and referential. The speaker literally told the audience that she was creating the black squares with her mind. Thus, we created an environment in which an intentional cue (the "intention-capture" flash) was provided, but where the speaker was neither looking nor pointing to the cued discourse element. In this situation, listeners did not have access to conventional cues about where the speaker's attention is directed, but if they were able to treat the intention-capture flash as referential, they would still have access to information about where the speaker meant to refer.

6.1. Methods

6.1.1. Participants

A total of 49 native speakers of English participated at the University of North Carolina, Chapel Hill, in exchange for course credit. Seven were excluded due to technical problems with the experimental program. This left 42 participants (21 in each of the 2 lists) in the final analysis.

6.1.2. Stimuli

The stimuli depicting Condition G and Condition H (attention-capture flash cues to N1 and N2, respectively) from Experiment 2 were used in this experiment. The critical difference in Experiment 3 was the presentation of an introductory video indicating that the attention-capture flash cue was referential and produced intentionally by the speaker. Thus, in Experiment 3, there were only two conditions:

Condition I – "Intention-capture" flash cue to N1.
Condition J – "Intention-capture" flash cue to N2.

Each subject again saw 58 videos in total. Although the same 34 fillers used in Experiment 1 were used again, an attention-capture flash was added to each, and always appeared over a character as it was mentioned, to encourage referential interpretations of the cue. These filler videos were then interspersed with the 24 target videos, 12 in each of the above conditions, with identity (e.g. Puppy vs. Panda Bear) and location of N1 (on the right- vs. the left-hand side of the screen) counterbalanced within each condition.

6.1.3. Procedure

Procedure was identical to that of Experiment 1, except that the experiment was preceded by a video, illustrated in Fig. 6, with the following content. This video can be viewed at http://arnoldlab.web.unc.edu/publications/supporting-materials/nappa-arnold/.

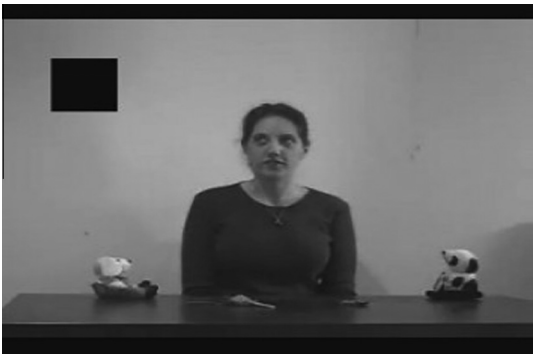

Fig. 6. Screen capture of first intention-capture cue in the introductory video used in Experiment 3. The speaker has just indicated the intentional nature of the cue by saying, "Want to see what I can do?"

"Hi, there! My name is Becky. Want to see what I can do? (flash in upper left of screen.) Did you see it? Watch, I'll do it again. (flash in upper left of screen.) Now I'll do it on the other side. (flash in upper right of screen.) It's really easy, watch this!

"Puppy (flash on Puppy) is talking to Panda Bear (flash on Panda Bear). Puppy (flash on Puppy) wants the candy bar (flash on candy bar). Now here comes Bunny and Froggy (Experimenter visually presents Bunny and Froggy). Bunny (flash on Bunny) wants the candy bar (flash on the candy bar) too.

"Now I'll tell you some other stories about these same characters. See which ones you like best!"

This introductory video was intended to make participants treat the attention-capture cue as intentional and referential. It was then followed by the 58 videos.

6.2. Results

Results demonstrate that participants used the new "intention-capture" cue referentially, as the order-of-mention bias was moderated by the location of the cue (see Fig. 7). When N1 was cued, participants seldom arrived at an N2 interpretation (~5% N2 interpretations), but the order-of-mention bias was eliminated when cues were to N2 (~45% N2 interpretations). These findings were verified by the mixed effects logistic regression,[3] which revealed a significant main effect of Cue Location (N1 vs. N2) ($\beta = -2.28$ (SE = 0.2); $t = -11.25$, $p < .0001$). This model also included the control variable List, which was also significant ($\beta = -1.26$ (SE = 0.36); $t = -3.54$, $p = 0.0013$). There were also shorter latencies for Intention-capture cues to N1 than cues to N2 ($\beta = -0.05$ (SE = 0.01); $t = -3.96$, $p < 0.001$).

6.3. Discussion

Results of Experiment 3 demonstrate, first and foremost, the importance of intentional information in pronoun resolution. Despite minimal exposure to the intention-capture cue (just the brief introductory video), participants successfully factored the cue into their interpretations of the ambiguous pronoun. This demonstrates that the null effect of the attention-capture cue in Experiment 2 was not simply a result of its unfamiliarity. By contrast, it suggests that intentional information, in and of itself, is quickly and successfully integrated into a referential interpretation. These findings support the Speaker-intention hypothesis.

At the same time, these results are also consistent with a variant of the Probabilistic Cues hypothesis. Notably, the intention-capture cue alone did not override order-of-mention biases. Rather, results of the intention-capture cue manipulation in Experiment 3 look quite similar to results of the gaze cue manipulation in Experiment 1 – either gaze OR intention-capture can compete with discourse biases,

[3] This model included a random slope for Cue Location by participants, but not by items.

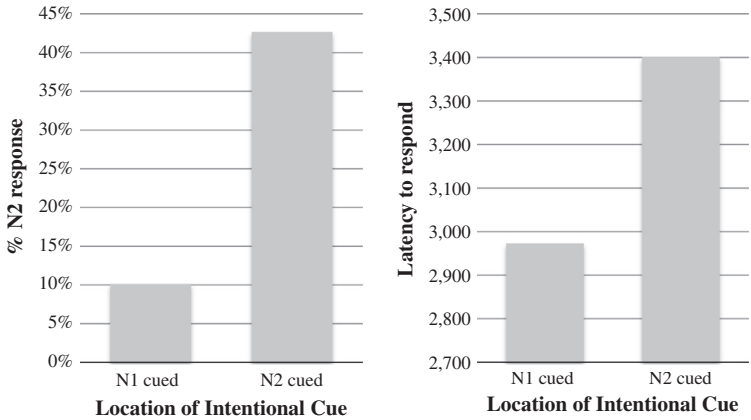

Fig. 7. Results of Experiment 3: proportion of N2 responses (left panel) and latencies to respond to the pronoun question (right panel) when "intention-capture" cues were to N1 (first-mentioned character in preceding sentence) and N2 (second-mentioned character in preceding sentence).

leading to a rough split of interpretations between N1 and N2. Thus, one interpretation of these results is that "more cues are better than one".

But critically, such a system must necessarily be restricted to information about the speaker's referential intentions, since the attention-capture cue itself in Experiment 2 had no effect. Moreover, Experiment 3 revealed that listeners are able to utilize even novel and recently-learned cues, like the intention-capture cue.

7. General discussion

The results of the three experiments here demonstrated that the interpretation of ambiguous pronouns is influenced by (1) the well-known first-mention bias, (2) evidence about the speaker's attentional focus via eyegaze, (3) evidence about the speaker's referential intentions from familiar pointing cues, and (4) evidence about the speaker's referential intentions from novel computerized "intention-capture" cues. By contrast, visual attentional capture cues only influenced participants' accuracy in visual memory, and not their interpretation of ambiguous pronouns.

These findings allow us to distinguish between several potential mechanisms of pronoun resolution. We first considered the Grammatical Constraints hypothesis. Under this view, discourse focus (as indicated by regularities such as the N1 bias) may be relevant as a linguistic category in its own right, independent from its relationship to representations about the speaker's intentions. Despite the fact that the term "focus" or "focus of attention" implies that attentional representations are related to discourse structure, it is possible that listeners use these categories directly, as a part of their linguistic knowledge about the appropriate use and interpretation of pronouns. The strong effects of pointing on pronoun resolution (Exp. 1) are consistent with the Grammatical Constraints view, since they could trigger a rule about deictic pronoun interpretation. However, the effects of both gaze (Exp. 1) and intention-capture (Exp. 3) are difficult to reconcile with this approach, or at least with a view that *only* grammatical rules matter.

We also considered the Probabilistic Cues hypothesis, which suggested that listeners learn to use any and all sources of information that indicate the correct pronoun interpretation. This hypothesis gains some support from the fact that pronoun resolution is indeed constrained by multiple cues, including gazing, pointing, and the novel intention-capture cue. However, the intention-capture effect demonstrated that listeners are able to quickly learn a novel and unfamiliar cue to the speaker's intentions, showing that prior familiarity is not a requirement for cue effectiveness. At the same time, the

identical cue was ineffective in Experiment 2, where it directed listeners' attention, but was not linked to the speaker's meaning. This suggests that all cues are not created equal. Instead, listeners appear to prioritize information that is directly tied to the speaker's referential meaning. Even though the listener's own attentional focus might have been used as a proxy for shared attention, Experiment 2 suggested that this is not a reliable mechanism for pronoun resolution.

Instead, the findings from these experiments support the Speaker's Intentions hypothesis. Referential assignment appears to largely be a process of inferring the objects or individuals to which the speaker *intended* to refer, rather than a process driven by the listener's own internal representations of the relative salience of elements in a discourse. The strongest effect we observed was pointing, which is a conventional gesture for indicating a referent. The intention-capture cue had a similar effect as gazing did. Both of these cues are likely to be interpreted as providing probabilistic, but still ambiguous information about the speaker's referential intentions. Speakers often look at the things they refer to, but not always. Likewise, the introduction video in Experiment 3 suggested that the speaker was "thinking about" the character that flashed, which could be taken to indicate reference, but not necessarily.

In this way, gazing and the intention-capture cue functioned much as discourse focus does. The first-mentioned character in an utterance is a fairly good indicator of the speaker's focus of attention, and is correlated with a tendency for the speaker to continue talking about that character in the next utterance (Arnold, 1998). Thus, first-mention status is a pretty good – but not perfect – signal of the speaker's attention (Arnold, 1998, 2010; Bock, 1986; Brennan, 1995). The fact that discourse status effects are modulated by other cues to the speaker's intentions suggests that these are not separate probabilistic cues, but rather a collection of information about the same thing: the speaker's intended referent.

Even if pronoun interpretation is fundamentally guided by inferences about the speaker's intentions, there is good reason to suspect that the listener's egocentric focus of attention might matter as well. Speakers don't typically use pronouns unless the listener can be expected to identify the referent on the basis of shared information. Thus, the most likely referent is something that is jointly attended, which in turn is most likely to be something that is co-present and/or is linguistically prominent. However, tracking the speaker's attention might be costly, and might lead listeners to use their own focus of attention as a readily available substitute. Nevertheless, the results of Experiment 2 argued against this: the listener's attention at pronoun onset facilitated visual memory, but had no effect on their interpretation of the pronoun.

This finding is notably different from evidence that listeners do appear to be influenced by their own egocentric focus of attention when building a representation of discourse accessibility. Arnold and Lao (submitted for publication) monitored participants' eye movements as they heard a story about two same-gender characters, e.g. *Birdy went apple-picking with Doggy near the farmhouse. He was wearing a hat,*…. At the ambiguous pronoun *he*, participants were relatively more likely to look at the first-mentioned character (*Birdy*), consistent with earlier work Arnold et al. (2000). But critically, this bias was relatively stronger when participants had also spent more time fixating Birdy in the first second of the trial, and relatively weaker when they had focused instead on Doggy in the first second of the trial. Thus, listeners' attention at the start of the trial modulated the strength of the first-mention bias – even when the listeners' attention was driven by an attention-capture cue that they were not aware of. Thus, Arnold & Lao's findings indicate a role for egocentric attention in building an initial prominence hierarchy as a scene is framed: elements that receive more attention in this early stage are considered somewhat more central to the discourse, even when the ensuing language focuses on another character. This account is consistent with other findings that demonstrate that early attention to a character in a scene increases the likelihood that it will be construed as the agent of an action, and used as a sentential subject to describe an ambiguous event (Gleitman et al., 2007; Hwang & Elsi, 2009).

In these cases, the effect of egocentric attention may be apparent because the other cues in the discourse are ambiguous about both discourse focus and the speaker's focus of attention. Order of mention is not a perfect indicator of discourse focus, and referential scenes can be described in numerous ways. Yet other features of the context can make one character more salient to all discourse participants, and therefore the listener's attention may be a partially reliable indicator of the character that is more likely to be referred to.

In contrast, Experiment 2 examined a different kind of attention: the listener's momentary visual attraction to one character in the scene, at the moment of hearing the pronoun. This moment follows several seconds in which the listener has already established a representation of the discourse context, which is likely to be more important for pronoun resolution. In natural communication situations, listeners are likely to fixate many objects in the environment, and not necessarily the ones that the speaker is currently referring to. In this situation, egocentric attention is not a reliable indicator of speaker meaning, and thus is ignored for the purpose of pronoun resolution.

The results of these three experiments, taken in sum, lend themselves to a model of referential resolution in which listeners integrate information to arrive at the most likely interpretation of the ambiguity, given the relative reliability and informativeness of each contributing cue. This is consistent with the idea that listeners use multiple cues probabilistically, depending on their utility. Crucially, though, "the most likely interpretation" must be interpreted as the most likely *intended meaning* (what the speaker was most likely aiming to communicate), rather than some simpler metric of likelihood constrained only by contingencies between linguistic units in the discourse, or even incorporating extra-linguistic contextual information. Although a great deal of prior research has emphasized the importance of treating ambiguity resolution across many linguistic domains as a probabilistic and constraint-based process, we wish to emphasize here the importance of building an intentional component into the assessments of maximum likelihood, reliability, and informativeness, at least in any such theory that pertains to assigning reference.

Acknowledgment

This research was partially supported by NSF Grant BCS-0745627 to J. Arnold.

Appendix A

A.1. Experimental stimuli

1. This story is about Puppy and Panda Bear. This is Puppy and this is Panda Bear. Puppy is having lunch with Panda Bear. He wants some milk.
2. This story is about Panda Bear and Puppy. This is Panda Bear and this is Puppy. Panda Bear is at school with Puppy. He wants the book.
3. This story is about Bunny and Froggy. This is Bunny and this is Froggy. Bunny is playing outside with Froggy. She wants the ball.
4. This story is about Froggy and Bunny. This is Froggy and this is Bunny. Froggy is getting ready for school with Bunny. She wants the toothbrush.
5. This story is about Puppy and Panda Bear. This is Puppy and this is Panda Bear. Puppy is eating dinner with Panda Bear. He wants some chicken.
6. This story is about Panda Bear and Puppy. This is Panda Bear and this is Puppy. Panda Bear is making dinner with Puppy. He wants the salt.
7. This story is about Bunny and Froggy. This is Bunny and this is Froggy. Bunny is washing dishes with Froggy. She wants the sponge.
8. This story is about Froggy and Bunny. This is Froggy and this is Bunny. Froggy is making art with Bunny. She wants the stapler.
9. This story is about Panda Bear and Puppy. This is Panda Bear and this is Puppy. Panda Bear is taking a bath with Puppy. He wants the soap.
10. This story is about Bunny and Froggy. This is Bunny and this is Froggy. Bunny is folding laundry with Froggy. She wants the blue towel.
11. This story is about Puppy and Panda Bear. This is Puppy and this is Panda Bear. Puppy is coloring with Panda Bear. He wants the green crayon.
12. This story is about Puppy and Panda Bear. This is Puppy and this is Panda Bear. Puppy is reading stories with Panda Bear. He wants the brown book.

13. This story is about Puppy and Panda Bear. This is Puppy and this is Panda Bear. Puppy is writing a story with Panda Bear. He wants the blue pen.
14. This story is about Puppy and Panda Bear. This is Puppy and this is Panda Bear. Puppy is getting ready for bed with Panda Bear. He wants the blanket.
15. This story is about Panda Bear and Puppy. This is Panda Bear and this is Puppy. Panda Bear is baking cookies with Puppy. He wants the bowl.
16. This story is about Panda Bear and Puppy. This is Panda Bear and this is Puppy. Panda Bear is having a snack with Puppy. He wants some chocolate milk.
17. This story is about Panda Bear and Puppy. This is Panda Bear and this is Puppy. Panda Bear is cleaning up with Puppy. He wants the broom.
18. This story is about Bunny and Froggy. This is Bunny and this is Froggy. Bunny is planting flowers with Froggy. She wants the seeds.
19. This story is about Panda Bear and Puppy. This is Panda Bear and this is Puppy. Panda Bear is having pizza with Puppy. He wants a pepperoni slice.
20. This story is about Froggy and Bunny. This is Froggy and this is Bunny. Froggy is playing with blocks with Bunny. She wants the blue block.
21. This story is about Froggy and Bunny. This is Froggy and this is Bunny. Froggy is making a sandwich with Bunny. She wants the bread.
22. This story is about Froggy and Bunny. This is Froggy and this is Bunny. Froggy is playing a board game with Bunny. She wants the blue guy.
23. This story is about Froggy and Bunny. This is Froggy and this is Bunny. Froggy is having a snack with Bunny. She wants some raisins.
24. This story is about Froggy and Bunny. This is Froggy and this is Bunny. Froggy is watching a movie with Bunny. She wants the popcorn.

A.2. Fillers

1. This story is about Puppy and Panda Bear. This is Puppy and this is Panda Bear. Puppy is playing in the sandbox with Panda Bear. He wants the shovel.
2. This story is about Froggy and Bunny. This is Froggy and this is Bunny. Froggy is playing cards with Bunny. Froggy wants the blue deck.
3. This story is about Puppy and Panda Bear. This is Puppy and this is Panda Bear. Puppy and Panda Bear are playing with toys together. Panda Bear wants the turtle.
4. This story is about Bunny and Froggy. This is Bunny and this is Froggy. Bunny and Froggy are going to watch a movie together. Froggy wants some popcorn.
5. This story is about Froggy and Bunny. This is Froggy and this is Bunny. Froggy and Bunny are playing with stickers together. Froggy wants the big one.
6. This story is about Puppy and Panda Bear. This is Puppy and this is Panda Bear. Puppy and Panda Bear are going to McDonalds. Puppy wants some McNuggets.
7. This story is about Bunny and Froggy. This is Bunny and this is Froggy. Bunny is playing outside with Froggy. Bunny wants to play with the snake.
8. This story is about Froggy and Bunny. This is Froggy and this is Bunny. Froggy is at the farm with Bunny. Froggy wants to play with the horse.
9. This story is about Puppy and Froggy. This is Puppy and this is Froggy. Puppy and Froggy are making breakfast together. Puppy wants some eggs.
10. This story is about Panda Bear and Bunny. This is Panda Bear and this is Bunny. Panda Bear and Bunny are playing with toys. Bunny wants the helicopter.
11. This story is about Froggy and Panda Bear. This is Froggy and this is Panda Bear. Froggy is playing play-doh with Panda Bear. Panda Bear wants to make something out of the blue play-doh.
12. This story is about Puppy and Panda Bear. This is Puppy and this is Panda Bear. Puppy is playing airport with Panda Bear. Puppy wants to play with the airplane first.
13. This story is about Puppy and Bunny. This is Puppy and this is Bunny. Puppy is playing dolls with Bunny. Bunny wants the blonde doll.

14. This story is about Panda Bear and Puppy. This is Puppy and this is Panda Bear. Panda Bear is having a drink with Puppy. Panda Bear wants the chocolate milk.
15. This story is about Bunny and Puppy. This is Bunny and this is Puppy. Bunny wants to have some milk with Puppy. Bunny wants the blue glass.
16. This story is about Puppy and Panda Bear. This is Puppy and this is Panda Bear. Puppy and Panda Bear are going to light some candles. Puppy wants to light the green candle.
17. This story is about Bunny and Froggy. This is Bunny and this is Froggy. Bunny and Froggy are playing with dolls. Froggy wants to put the doll in the bassinette.
18. This story is about Froggy and Panda Bear. This is Froggy and this is Panda Bear. Froggy and Panda Bear want to go fishing. Froggy wants the fishing rod.
19. This story is about Panda Bear and Bunny. This is Panda Bear and this is Bunny. Panda Bear and Bunny are getting ready to go outside. Bunny wants to put on the red hat.
20. This story is about Froggy and Puppy. This is Froggy and this is Puppy. Froggy and Puppy are going to go play outside. she wants to wear the white shoes.
21. This story is about Panda Bear and Puppy. This is Panda Bear and this is Puppy. Panda Bear and Puppy are getting ready for school. Panda Bear wants to wear the jacket.
22. This story is about Froggy and Panda Bear. This is Froggy and this is Panda Bear. Froggy and Panda Bear are playing with animals. Froggy wants to play with the turtle.
23. This story is about Panda Bear and Puppy. This is Pand Bear and this is Puppy. Panda Bear and Puppy are going to go play outside in the water. Panda Bear wants the squirt gun.
24. This story is about Puppy and Froggy. This is Puppy and this is Froggy. Puppy is putting together a puzzle with Froggy. Puppy wants the purple piece.
25. This story is about Froggy and Puppy. This is Froggy and this is Puppy. Froggy wants to watch a movie with Puppy. Puppy wants the remote control.
26. This story is about Bunny and Froggy. This is Bunny and this is Froggy. Bunny wants to go shopping with Froggy. Bunny wants the dollar bill.
27. This story is about Puppy and Panda Bear. This is Puppy and this is Panda Bear. Puppy and Panda Bear are going to take some pictures together. Puppy wants to hold the camera first.
28. This story is about Froggy and Panda Bear. This is Froggy and this is Panda Bear. Froggy and Panda Bear have to wash dishes together. Froggy does not want to wash the glasses.
29. This story is about Puppy and Bunny. This is Puppy and this is Bunny. Puppy is going to make some art with Bunny. Bunny wants the play-doh.
30. This story is about Bunny and Panda Bear. This is Bunny and this is Panda Bear. Bunny wants to race boats with Panda Bear. Bunny wants to be the green boat.
31. This story is about Panda Bear and Puppy. This is Panda Bear and this is Puppy. Panda Bear and Puppy are going to dress up the doll. Panda Bear wants her to wear the pink shirt.
32. This story is about Bunny and Froggy. This is Bunny and this is Froggy. Bunny is getting ready for a party with Froggy. Bunny wants to blow up more balloons.
33. This story is about Panda Bear and Bunny. This is Panda Bear and this is Bunny. Panda Bear is going to put stickers in the sticker book with Bunny. Panda Bear wants the sparkly sticker.
34. This story is about Panda Bear and Puppy. This is Panda Bear and this is Puppy. Panda Bear and Puppy are going to race cars. Puppy wants to be the red car.

References

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

Ariel, Mira. (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. (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*(4), 187–203.

Arnold, J. E., & Lao, S. C. (submitted for publication). Effects of psychological attention on pronoun comprehension. Ms., University of North Carolina at Chapel Hill.
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., Kaiser, E., Kahn, J., & Kim, L. (2013). Information structure; Linguistic, cognitive, and processing approaches. *WIREs Cognitive Science*. Wiley, http://dx.doi.org/10.1002/wcs.1234.
Baldwin, D. A. (1995). Understanding the link between joint attention and language. In C. Moore, & P. J. Dunham (Eds.), *Joint attention: Its origins and role in development*.
Bock, J. K. (1986). Meaning, sound, and syntax: Lexical priming in sentence production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12*, 575–586.
Brennan, S. E. (1995). Centering attention in discourse. *Language and Cognitive Processes, 102*, 137–167.
Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory and Cognition, 22*, 482–1493.
Brennan, S. E., Friedman, M. W., & Pollard, C. J. (1987). A centering approach to pronouns. In *Proceedings, 25th annual meeting of the ACL* (pp. 155–162). Stanford, CA: Association of Computational Linguistics.
Carpenter, M., Nagell, K., & Tomasello, M. (1998). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monographs of the Society for Research in Child Development, 63*(4).
Chafe, W. L. (1994). *Discourse, consciousness, and time*. Chicago: Chicago University Press.
Clark, H. H., & Krych, M. A. (2004). Speaking while monitoring addressees for understanding. *Journal of Memory and Language, 50*, 62–81.
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.
Clark, H. H., & Sengul, C. J. (1979). In search of referents for nouns and pronouns. *Memory & Cognition, 7*, 35–41.
Fukumura, K., & Van Gompel, R. P. G. (2010). Choosing anaphoric expressions: Do people take into account likelihood of reference? *Journal of Memory and Language, 62*, 52–66.
Georgiades, M. S., & Harris, J. P. (1997). Biasing effects in ambiguous figures: Removal or fixation of critical features can affect perception. *Visual Cognition, 4*, 383–408.
Gernsbacher, M. A. (1989). Mechanisms that improve referential access. *Cognition, 32*, 99–156.
Gernsbacher, M. A., & Hargreaves, D. J. (1988). Accessing sentence participants: The advantage of first mention. *Journal of Memory and Language, 27*, 699–717.
Gerrig, R. J., Horton, W. S., & Stent, A. (2011). Production and comprehension of unheralded pronouns: A corpus analysis. *Discourse Processes, 48*, 161–182.
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, W., & Hudson Kam, C. L. (2009). Co-speech gesture as input in verb learning. *Developmental Science, 12*, 81–87.
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.
Gordon, P. C., Grosz, B. J., & Gilliom, L. A. (1993). Pronouns, names, and the centering of attention in discourse. *Cognitive Science, 17*, 311–347.
Greene, S. B., McKoon, G., Gerrig, R. J., & Ratcliff, R. (1994). Unheralded pronouns and management by common ground. *Journal of Memory and Language, 33*, 511–526.
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., & 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.
Horton, W. S., & Keysar, B. (1996). When do speakers take into account common ground? *Cognition, 59*, 91–117.
Hudson-D'Zmura, S., & Tanenhaus, M. K. (1998). Assigning antecedents to ambiguous pronouns: the role of the center of attention as the default assignment. In M. Walker, A. Joshi, & E. Prince (Eds.), *Centeringtheory in discourse* (pp. 199–226). Oxford: Oxford University Press.
Hwang, H., & Elsi, K. (2009). The effects of lexical vs. perceptual primes on sentence production in Korean: An online investigation of event apprehension and sentence formulation. *Talk presented at the 22nd annual meeting of the CUNY conference on human sentence processing, UC Davis, March 2009*.
Kehler, A. (1997). Current theories of centering for pronoun interpretation: A critical evaluation. *Computational Linguistics (Squibs and Discussions)*, 23.
Kehler, A., Kertz, L., Rohde, H., & Elman, J. L. (2008). Coherence and coreference revisited. *Journal of Semantics, 25*, 1–44.
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.
Langton, S. R. H., & Bruce, V. (2000). You must see the point: Automatic processing of cues to the direction of social attention. *Journal of Experimental Psychology: Human Perception and Performance, 26*, 747–757.
Pickering, M., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences, 27*(2), 169–190.

Pitz, G. (2007). *Behavior and Cognitive Research (BCR)*. Software program, University of North Carolina, Chapel Hill.
Salverda, A. P., & Altmann, G. T. M. (2011). Attentional capture of objects referred to by spoken language. *Journal of Experimental Psychology: Human Perception and Performance, 37*, 1122–1133.
Sanford, A. J., & Garrod, S. C. (1981). *Understanding written language*. Chichester: Wiley.
Stevenson, R. J., Crawley, R. A., & Kleinman, D. (1994). Thematic roles, focus and the representation of events. *Language and Cognitive Processes, 9*, 473–592.
Tetrault, J. (2001). A corpus-based evaluation of centering and pronoun resolution. *Computational Linguistics, 27*, 507–520.