Cognition 182 (2019) 275–285

Contents lists available at ScienceDirect

Cognition

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

Original Articles

What do you know? ERP evidence for immediate use of common ground during online reference resolution

Les Sikos[a,b,*], Samuel B. Tomlinson[b], Conor Heins[b], Daniel J. Grodner[b]

[a] *Psycholinguistics Group, Department of Language Science and Technology (FR 4.7), Saarland University, 66123 Saarbrücken, Germany*
[b] *Department of Psychology, Swarthmore College, 500 College Avenue, Swarthmore, PA 19081, USA*

ARTICLE INFO

Keywords:
Language comprehension
Perspective taking
Common ground
Referential ambiguity
Visual world
Nref effect

ABSTRACT

Recent evidence on the time-course of conversational perspective taking is mixed. Some results suggest that listeners rapidly incorporate an interlocutor's knowledge during comprehension, while other findings suggest that listeners initially interpret language egocentrically. A key finding in support of the egocentric view comes from visual-world eye-tracking studies — listeners systematically look at potential referents that are known to them but unknown to the speaker. An alternative explanation is that eye movements might be driven by attentional processes that are unrelated to referent identification. To address this question, we assessed the time-course of perspective taking using event-related potentials (ERP). Participants were instructed to select a referent from a display of four animals (e.g., "Click on the brontosaurus with the boots") by a speaker who could only see three of the animals. A competitor (e.g., a brontosaurus with a purse) was either mutually visible, visible only to the listener, or absent from the display. Results showed that only the mutually visible competitor elicited an ERP signature of referential ambiguity. Critically, ERPs exhibited no evidence of referential confusion when the listener had privileged access to the competitor. Contra the egocentric hypothesis, this pattern of results indicates that listeners did not consider privileged competitors to be candidates for reference. These findings are consistent with theories of language processing that allow socio-pragmatic information to rapidly influence online language comprehension. The results also suggest that eye-tracking evidence in studies of online reference resolution may include distraction effects driven by privileged competitors and highlight the importance of using multiple measures to investigate perspective use.

1. Introduction

Interlocutors make social inferences when faced with ambiguous linguistic utterances. For example, if a speaker says, "May I borrow that book?", the listener must utilize information from the social context to constrain the set of possible referents (all books) to the most likely candidate before initiating a response. One way that listeners may understand ambiguous utterances is by maintaining a representation of the *common ground* (i.e., information that is mutually shared between interlocutors) and the *privileged ground* (i.e., information that is privileged to either the speaker or listener). Such a representation allows listeners to make rapid, context-sensitive inferences about the speaker's intentions. For instance, in one context, the listener may utilize common ground information to infer that the referent of "that book" is a mutually salient object (e.g., a book that both interlocutors can see) rather than one that is known only to the listener (e.g., a book in the listener's backpack). Distinguishing between common and privileged information requires that individuals keep track of what people around them know, and how this differs from their own knowledge. This process of *perspective taking* is a fundamental and ubiquitous form of social cognition and is necessary for interpreting virtually every linguistic utterance (Clark & Carlson, 1981; Clark & Marshall, 1981; Clark, 1996).

Researchers widely agree that listeners ultimately use their mental representation of the speaker's knowledge to understand an utterance. However, it is unclear how rapidly and fully listeners can access ground information during online processing (for reviews see Barr & Keysar, 2006; Brennan & Hanna, 2009; Brown-Schmidt & Hanna, 2011; Brown-Schmidt & Heller, 2014). The vast majority of experimental studies investigating this question have utilized visual-world eye-tracking paradigms in which a participant listener responds to commands that require interacting with a display of objects, either in the real world or on a computer screen, while their eye movements are recorded (e.g., Eberhard, Spivey-Knowlton, Sedivy, & Tanenhaus, 1995). Crucially, some objects are hidden from the speaker, thereby creating differences

---

[*] Corresponding author at: Psycholinguistics Group, Department of Language Science and Technology, Saarland University, 66123 Saarbrücken, Germany.
E-mail address: sikos@coli.uni-saarland.de (L. Sikos).

https://doi.org/10.1016/j.cognition.2018.10.013
Received 12 April 2017; Received in revised form 8 October 2018; Accepted 14 October 2018

in perspective between the speaker and the listener. Behavioral and eye movement evidence from such studies regarding the time course at which ground information can be used to resolve a referent has been equivocal. Some findings have shown that listeners are more likely to look at a privileged object (e.g., a book visible only to the listener) when it is a competitor to the target in common ground (e.g., a mutually visible book) than when the privileged object is not a competitor (e.g., a cup visible only to the listener), and that they are delayed in ultimately selecting the correct object (Apperly et al., 2010; Barr, 2008; Keysar, Barr, Balin, & Brauner, 2000; Keysar, Barr, Balin, & Paek, 1998; Keysar, Lin, & Barr, 2003; Wu & Keysar, 2007). These findings suggest that listeners are slow to take the speaker's perspective into consideration. In contrast, other studies have argued that referents determined by common ground considerations are favored from the earliest stages of referent resolution (Brown-Schmidt, 2009; Hanna & Tanenhaus, 2004; Hanna, Tanenhaus, & Trueswell, 2003; Heller, Grodner, & Tanenhaus, 2008). This second set of results suggests that listeners immediately incorporate the speaker's perspective.

Multiple theories have been proposed for how listeners might track a speaker's perspective during online reference resolution. One school of thought posits that listeners are inherently *egocentric* — that is, they initially utilize information available to the self and only later adjust their interpretation to account for the speaker's perspective (Keysar et al., 1998). Thus, as a referential description unfolds, a listener's first-pass interpretation would consider all matching referents in their egocentric perspective as potential candidates. For instance, in the example above, the listener may initially consider both the book in her backpack, as well as the one that is mutually visible, as potential referents, because both objects match the referring expression "that book". These egocentric accounts are motivated by considerations of cognitive efficiency. Because the speaker's knowledge is not directly available to the listener, is potentially open ended, and may even be in direct conflict with the listener's perceptual knowledge and attention, the listener may find it cognitively effortful to consider such information.

At least two variants of the egocentric hypothesis have been put forth. The *Perspective-Adjustment* model holds that first-pass interpretation is entirely egocentric (Keysar et al., 1998, 2000). Consequently, taking the perspective of the speaker necessarily requires optional, costly, and time-consuming secondary processes such as inhibiting the egocentric information activated during the initial stage. Although this hypothesis is consistent with evidence showing that listeners can be significantly delayed in taking the speaker's perspective into consideration (Keysar et al., 1998, 2000, 2003; Wu & Keysar, 2007), it cannot explain findings showing early bias toward common ground referents over privileged competitors (Hanna & Tanenhaus, 2004; Hanna et al., 2003; Heller et al., 2008).

A second, more nuanced version of the egocentric hypothesis is the *Anticipation-Integration* model (Barr, 2008, 2014, 2016). It holds that perceivers can strategically use common ground information to make top-down predictions about potential referents prior to hearing a referring expression (the anticipation phase). However, common ground information is completely ignored while the referring expression is being processed (the integration phase) (Barr, 2008, 2014). This view claims early fixation biases to common ground referents are a product of anticipation processes. Importantly, top-down anticipation is proposed to be informationally-encapsulated from bottom-up processes during the integration of the referring expression with candidate referents. As a result, despite early orienting to the common ground, all referents that are compatible with the unfolding description are predicted to receive equal activation from the integration process. In support of this model, there is evidence that fixations to both common ground referents and privileged referents increase at the same rate as the description is processed (Barr, 2008, 2016).

An alternative to these egocentric accounts is the *constraint-based hypothesis.* This account claims that multiple probabilistic constraints interact to guide reference resolution, and each can exert its influence as it becomes available (Brown-Schmidt & Hanna, 2011; Hanna et al., 2003; Heller et al., 2008). Thus, rather than being inherently egocentric, listeners constantly weigh both social cues (e.g., ground information) and non-social cues (e.g., perceptual information) based on various factors, such as the nature of conversational exchange, the goals of the exchange, the types of cues that are available, and so on. On this account, ground information can in principle be integrated as soon as it becomes available. However, because perceptual information is also taken into consideration, any objects in privileged ground that share perceptual features with the target should lead to some amount of interference. A critical prediction of this hypothesis is that the amount of interference from a competitor object should be modulated by ground information, such that privileged competitors will lead to less interference than shared competitors. This account is also able to explain many of the contradictory findings in previous work. The ability of the system to make immediate use of ground information is compatible with results showing early effects of perspective taking (Brown-Schmidt, 2009; Hanna & Tanenhaus, 2004; Hanna et al., 2003; Heller et al., 2008; Nadig & Sedivy, 2002), and the interference from perceptually similar objects explains why the eyes are drawn to competitor objects that are in privileged ground (Keysar et al., 1998, 2000, 2003; Wu & Keysar, 2007).

This final point raises an important question about the linking hypothesis between eye-tracking results in visual-world studies of perspective taking and the conclusions that are ultimately drawn. A critical assumption of all the studies discussed above is that eye movements in this paradigm are predominantly driven by referential processing. That is, participants' eye movements are assumed to indicate which objects are being entertained as potential referents for a referring expression. This assumption may be too strong.

An alternate possibility is that eye movements and reaction times in visual world studies may be influenced by additional factors beyond mapping the referential description onto the immediate visual environment. Previous work has shown that eye movements can be driven by the degree of phonological overlap between the acoustic input and the phonological forms of potential referents in the display (Allopenna, Magnuson, & Tanenhaus, 1998; Dahan, Magnuson, Tanenhaus, & Hogan, 2001), and can also reflect the activation of semantic information related to the acoustic input. For example, Yee and Sedivy (2006) found that when participants hear a word like "lock", their eye movements are drawn to images of both the named object and semantically related objects (e.g., key). In the present paradigm, the privileged competitor is always phonologically and semantically related to the target. Thus, these low-level factors should attract the perceiver's attention away from the target independent of referent identification. We will refer to these non-referential effects as *attentional distraction*.

Attentional distraction need not be low-level. Perspective use does not eliminate the importance of the privileged competitor, it merely eliminates it as a candidate referent. For instance, the perceiver's attention may be drawn to the privileged competitor not because they think they should select it, but because it could be relevant to their interlocutor. Someone who is interested in the book in your hands may also want to know about a related book concealed in your backpack. This does not render the referent of "Can I borrow that book?" ambiguous, but the request could still cause the perceiver to attend to their own privileged book.

In short, we cannot be sure why participants are slower to select a target and are more likely to fixate a privileged object when it is a competitor. It may be that the comprehension system truly entertains the competitor as a candidate for reference or it could be that the competitor merely draws attention from the target. Eye-tracking and other behavioral measures cannot distinguish between these competing explanations. The current study uses event-related brain potentials (ERP) to address this question.

1.1. Event-related potentials and referential ambiguity

Previous work using ERP has shown that the neurophysiological response to referentially ambiguous words differs from the response to unambiguous control words (Nieuwland & Van Berkum, 2008a, 2008b; Nieuwland, Otten, and Van Berkum, 2007; Van Berkum, Brown, & Hagoort, 1999; Van Berkum, Brown, Hagoort, & Zwitserlood, 2003; Van Berkum, Koornneef, Otten, & Nieuwland, 2007). For example, in (1a), the underlined word can refer to multiple equally suitable referents. This ambiguity elicits a sustained negative shift of the ERP relative to a context with one suitable referent (1b).

1a. There was a boy upstairs and a boy downstairs. The boy…
1b. There was a boy upstairs and a girl downstairs. The boy…

This brain response to referential ambiguity, or *Nref effect*, typically arises in the ERP signal 300–400 ms after the onset of the ambiguous word and is broadly distributed but dominant at frontal areas. This effect is robust across written and spoken stimuli, and across different descriptive forms (pronouns and full noun phrases) (see Van Berkum et al., 2007 for a review). Crucially, it is reliably elicited by competition between *viable* referential candidates. No Nref effect occurs when two referents match the referential description in the discourse, but only one of them is a potential candidate (e.g., if one of the boys in (1a) left the scene) (Nieuwland et al., 2007). Importantly, this negativity is extremely long lasting — it is sustained for 1000–2000 ms and has even been shown to persist for over 900 ms after the triggering ambiguity has been resolved (Nieuwland et al., 2007). The Nref effect can therefore be used as a diagnostic of even temporary referential ambiguities.

1.2. Current study

The present study leverages the neural response to referentially-ambiguous utterances to assess the time course of perspective use. Though ERP methods have been applied to investigate the time course of referent identification (Brodbeck, Gwilliams & Pylkkänen, 2015), no previous studies have attempted to use ERPs to investigate online perspective taking. We recorded participants' brain activity while they listened to auditory instructions from a *director* to click on a target (e.g., "… the brontosaurus with the boots") (Fig. 1). This target was always in common ground with the director. On critical trials, the display also contained an object that was a temporary competitor to the target (e.g., a brontosaurus with a purse), but which was either concealed from the director (Privileged Ground Competitor, PGC) or in common ground (Common Ground Competitor, CGC). We compared the ERPs elicited by these conditions to an unambiguous control condition in which there was no competitor (No Competitor, NoC) in the display.

If ERPs are sensitive to referential ambiguity in this paradigm, then the critical word (e.g., "brontosaurus") should elicit a reliable Nref effect in the CGC condition relative to the NoC condition. This is because the referent of "brontosaurus" is temporarily ambiguous between two candidates in common ground in the former case, but not the latter. The critical question was whether the PGC condition would also elicit an Nref effect. If so, this would indicate that participants consider the privileged competitor as a candidate for reference. Such a result would provide strong evidence in favor of the egocentric hypothesis. On the other hand, if the PGC condition patterns like the NoC condition, it would indicate that participants are not egocentric, and are instead able to quickly use ground information to restrict potential referents to those appearing in common ground.

To make these predictions about the time-course of perspective taking more concrete, consider Fig. 2, which provides an idealized depiction of referential processing in the current paradigm. Panel A shows that in the ambiguous CGC condition, referential ambiguity arises in the linguistic signal (A_L) at the auditory onset of the target noun ("brontosaurus") and lasts until disambiguating information becomes available (D_L). However, the brain's recognition of referential ambiguity (A_R) and its eventual resolution (D_R) are both likely to be delayed slightly, relative to when the information becomes available in the linguistic signal, partly because the transmission of information from the peripheral nervous system to higher-level cognitive processing areas requires time. Similarly, there are likely to be additional delays before the neurophysiological effects of referential ambiguity (A_E) and its subsequent resolution (D_E) become observable in the ERP signal as an Nref effect. It is currently unclear exactly which factors (e.g., word length, word frequency, task complexity, etc.) contribute to the total delay d between A_L and A_E. However, previous work indicates that there is considerable variability across studies, and this delay can be "much longer" than the typical 300–400 ms (Van Berkum, 2009).

The critical timing question the current study was designed to test is when ground information can be used to constrain potential candidates for reference. Panel B of Fig. 2 illustrates the egocentric account's prediction (red) for the PGC condition. On this account, listeners initially consider both brontosauri as potential candidate referents because the use of ground information is delayed until an effortful second stage of processing (G_{EGO}). Consequently, a meaningful period of referential ambiguity should exist on this account. For instance, Barr (2008, 2016) argues that PG competitors are on equal footing with the CG target as the referential description is being processed. Neither of these accounts explicitly states how long this equivalent activation might last. As an upper bound, Kronmueller et al. (2017) claim that certain types of ground information are delayed by 1000 ms after the onset of the referential description. Barr (2016) reviews eye movement studies with a design more like the present one. Based on that data, the PG competitor appears to compete with the CG target for *at least* 400–500 ms. Previous ERP work indicates that ambiguities of this duration elicit robust and long-lasting Nref effects (Boudewyn et al., 2015; Nieuwland & Van Berkum, 2006; Nieuwland & Van Berkum, 2008a, 2008b; Van Berkum et al., 1999, 2003). In fact, we can find no example in the literature where an Nref effect is shorter in duration than 950 ms, even in cases where the ambiguity has been resolved much earlier (Nieuwland et al., 2007). Thus, egocentric accounts predict that an Nref effect should be robustly observable in the PGC condition at some point during the ERP record.

Panel C of Fig. 2 illustrates the constraint-based account's prediction for the PGC condition. In contrast to above, this account argues that ground information can be used without delay to constrain potential candidate referents as soon as the information becomes available. In the current paradigm, ground information in the PGC condition should effectively be available immediately upon hearing the target word (G_{CBH}), because the visual scene is co-present. Thus, the constraint-based hypothesis predicts that the immediate use of ground information would effectively preclude referential ambiguity from arising at all (Ø). Consequently, no Nref effect should be triggered (Ø). Note that this is conceptually similar to the logic of Nieuwland et al. (2007), which used linguistic stimuli in which one of two potential candidate referents (e.g., boys) leaves the scene during the prior discourse, making a subsequent referential expression (e.g., "the boy") unambiguous.

In sum, if the PCG condition elicits an Nref effect at any point during the recorded ERP epoch, it would indicate that participants initially consider the privileged competitor to be a candidate for reference and only bring ground information to bear in order to resolve the ambiguity. In contrast, if no Nref effect is found, it would provide strong evidence that ground information is used immediately to constrain potential referents to only those candidates that are in common ground, essentially precluding an ambiguity from arising.

Finally, if the behavioral patterns of eye movements and reaction times reported in previous studies are confounded with attentional distraction effects, then we should also find evidence of such attentional distraction on response times in the current task. More specifically, if the privileged competitor draws *non-referential* attention away from the target, then we should find that listeners are delayed in ultimately

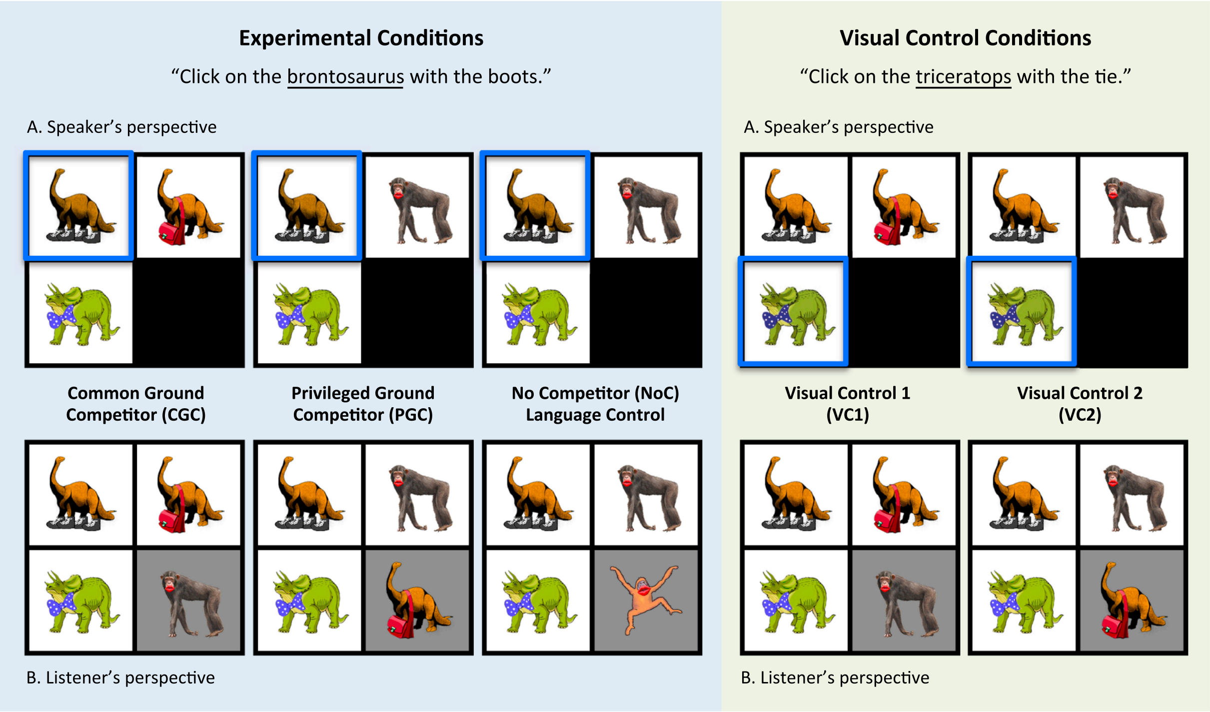

Fig. 1. Example Stimuli. *Left panel*: Experimental conditions. *Right panel:* Visual control conditions. Auditory stimulus (target word underlined) and corresponding visual stimuli, from the perspective of the speaker/director (above) and the listener/matcher (below). A white background indicates common ground. A gray background indicates privileged ground.

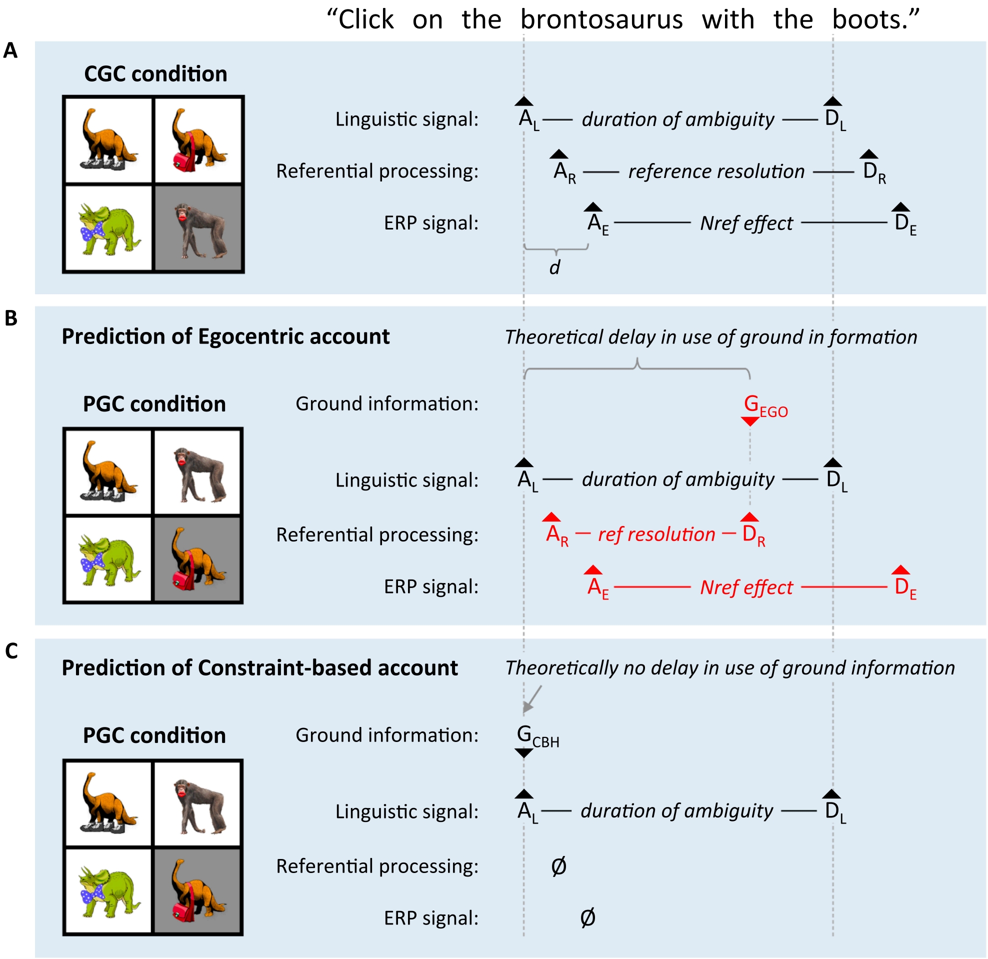

Fig. 2. An idealized depiction of referential processing in the current paradigm. (A) CGC condition. Referential ambiguity arises in the linguistic signal (A_L) at the auditory onset of the target noun ("brontosaurus") and lasts until disambiguating information becomes available (D_L). The brain's recognition of referential ambiguity (A_R) and resolution (D_R) are both delayed slightly, relative to when information becomes available in the linguistic signal. Neurophysiological effects of referential ambiguity (A_E) and resolution (D_E) are also delayed before becoming observable in the ERP signal as an Nref effect. It is currently unclear exactly which factors contribute to the total delay d between A_L and A_E. (B) The egocentric account's prediction (red) for the PGC condition. On this account, listeners initially consider both brontosauri as potential candidate referents because the use of ground information is delayed until a second stage of processing (G_{EGO}). Consequently, a meaningful period of referential ambiguity should exist on this account. If so, an Nref effect should be robustly observable at some point during the ERP record. (C) The constraint-based account's prediction for the PGC condition. This account argues that ground information can be used without delay to constrain potential candidate referents immediately upon hearing the target word (G_{CBH}), effectively precluding an ambiguity from arising (Ø). Consequently, no Nref effect should be triggered (Ø). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

selecting the correct target object in the PGC condition, while no Nref effect should be elicited.

2. Methods

2.1. Participants

Behavioral and electrophysiological data are reported for 50 right-handed students from Swarthmore College (18–22 years, $M = 18.98$, 26 males). All participants were native English speakers, had normal or corrected-to-normal vision, and had no history of neurological or psychiatric conditions. Data from eight additional participants were excluded due to excessive EEG artifacts (4), response accuracy below 90% (2), or data corruption (2). Participants received $20 or course credit.

2.2. Materials

Participants listened and responded to pre-recorded commands while viewing cartoon displays depicting a 2x2 grid containing four animals wearing accessory items. For each trial, three of the four animals were in front of a white background, indicating that they were visible to the director. The fourth animal was in front of a gray background, indicating that it was hidden from the director. The locations of the privileged quadrant and the animal-quadrant assignments were randomized. Visual stimuli were adapted from Brown-Schmidt (2009) by combining artwork of eleven polysyllabic[1] animals (alligator, brontosaurus, caterpillar, chimpanzee, hippopotamus, mountain lion, orangutan, rhinoceros, praying mantis, salamander, triceratops) and ten accessories (boots, cap, crown, flower, glasses, heels, lips, party hat, purse, tie). The initial phoneme of each animal and each accessory was distinct so that the regions of ambiguity and disambiguation would have clear onsets.

120 stimulus sets of three conditions each were created (Fig. 1, left panel) for each auditory command (e.g., “Click on the brontosaurus with the boots”). The CGC condition created a temporary referential ambiguity in the auditory stimulus at the animal name (“brontosaurus”) that was resolved by the accessory label (“boots”). The PGC displays were identical except that the competitor was in privileged ground. In this case, the auditory stimulus could be considered temporarily ambiguous at the target word only to the extent that participants considered the PGC to be a candidate for reference. The NoC condition was identical to the PGC condition except that the competitor was replaced by a different animal, thus making the animal referent unambiguous. Two additional conditions of 40 trials each served as visual controls (Fig. 1, right panel). These conditions were visually identical to the CGC and PGC displays, but the target was a singleton in common ground. These conditions ensured that participants could not predict the identity of the target animal in advance on the basis of the displays.

Three stimulus lists were created from the above visual display materials according to a Latin square design. The CGC, PGC, and NoC displays were counterbalanced across lists, while the secondary visual control conditions (VC1, VC2) did not vary across lists. Thus, each list contained 200 trials, with 20% containing a CG competitor, 20% containing a PG competitor, and 60% containing no competitor. Presentation order was pseudorandomly mixed such that no more than two trials of the same condition occurred in sequence, no more than three consecutive trials shared the same animal target, and each animal + accessory target appeared no more than once per condition.

Auditory stimuli were recorded with a natural speaking rate and intonation while two research assistants played a full 200-trial session using List 1 and a procedure identical to the first practice session described below. During the recording, the director was blind to the identity of the matcher’s privileged item. The director’s actual knowledge has been shown to be an important factor in appropriate perspective use (Heller, Grodner, & Tanenhaus, 2009). Audio files containing obtrusive speech disfluencies were re-recorded. Files with minor disfluencies were maintained in the final sample to increase the naturalness of the task. The audio files from List 1 were then applied to Lists 2 and 3. The order of audio files was held constant across lists such that for each critical item, the audio file was mapped to a different condition (CGC, PGC, NoC) in each list. On average the target noun began 2882 ms ($SD = 200$) after the command onset and had a mean duration of 652 ms ($SD = 29$). The disambiguating accessory began 879 ms ($SD = 112$) later. Three additional lists were created by reversing the original lists.

2.3. Procedure

Participants completed a survey[2] and two brief practice sessions before the main task. The first practice session familiarized the participant with director task (Keysar et al., 2000; Nadig & Sedivy, 2002), and the second introduced modifications designed to minimize eye movement and motor-response artifacts during EEG recording. During the initial practice session participants performed the basic task with an experimenter, first as the director (20 trials) and then as the matcher (20 trials). The director and matcher sat across from one another at separate monitors so they could not see each other’s displays. At the beginning of each trial the target in the director’s display was highlighted in blue. The matcher selected the intended referent using a game controller. At the end of this practice the participant was asked to explain what the director could see when he or she gave a command. If the participant did not provide an accurate description, an explicit description was provided.

Participants were then relocated to the EEG recording room and electrodes were applied. They sat approximately 100 cm from a display and external loudspeakers. Participants played the role of the matcher and followed the commands of a pre-recorded director. Participants were (truthfully) informed that this director did not know the identity of hidden objects during the recording. Each trial began with a 4600 ms display preview period in which participants were allowed to view the screen naturally. During this time a 2x2 grid consisting of black squares subtending 9° x 9° of visual angle appeared. After 600 ms the black squares were removed one at a time every 1000 ms revealing each animal. 1000 ms after the final animal, a bell sounded and a red fixation cross appeared. Participants were asked to pay close attention to the animals and accessories during the preview period, but to fix their gaze on the central cross upon hearing the bell and to not blink for the remainder of the trial. After a random duration (600–1000 ms) the director’s command began. To avoid motor responses contaminating the ERP signal, participants were instructed to delay responding until they heard a second bell at the end of the audio file. If responses preceded the second bell, the words “Too Fast” appeared on the screen indicating that the participant responded too quickly, and the trial was excluded from analyses. Trials in which the response followed the second bell were labeled “good trials.” The second practice session concluded once

---

[1] Polysyllabic words were used in order to increase the duration of any potential referential ambiguity.

[2] A secondary goal of the current study was to assess whether individual differences in social aptitude can modulate referential processing. We investigated whether people with strong social skills are better at taking perspective than people whose social skills are less strong. Social skill was operationalized using the ‘Social’ subscale of the Autism Spectrum Quotient (ASQ) survey, which measures the degree to which typical adults exhibit traits associated with Autism Spectrum Disorder (Baron-Cohen et al., 2001). Previous work from our lab has shown that these scores are correlated with individual differences in perspective taking during both comprehension and production (Grodner, Dalini, Pearlstein-Levy & Ward, 2012; Grodner, Cheek, & Hsieh, 2018). However, only minor effects of social aptitude were found and are therefore not reported.

10 consecutive good trials had been completed. This typically happened within 20 practice trials or less. After the practice session, the experimental trials began. Following the ERP portion, participants were asked to describe how they interpreted the gray square, whether they considered the director's display, and if they developed any strategies for making their choices. The entire session lasted approximately two hours.

2.4. EEG recording and analysis

Continuous EEG was recorded with a 64-channel HydroCel Geodesic Sensor Net™ and amplified with a DC-coupled high input impedance amplifier with onboard 400 Hz anti-aliasing hardware filter (200 MΩ, Net Amps 300™, Electrical Geodesics Inc. (EGI), Eugene, OR). Online voltages were referenced to the vertex (Cz), and amplified analog voltages were digitized at 1000 Hz. Vertical eye movements and blinks were monitored with electrodes above and below the eyes, and horizontal eye movements were monitored by electrodes at the outer canthi of each eye. Impedances were maintained below 40 kΩ prior to each block. Further off-line processing and analysis of the EEG signal was performed using a combination of NetStation 4.5.4 (EGI), EEGLAB (Delorme & Makeig, 2004), and ERPLAB (Lopez-Calderon & Luck, 2014) toolboxes. Prior to segmentation, data were down-sampled to 200 Hz and bandpass filtered (0.03–40 Hz). For each trial, EEG was segmented into epochs spanning -200 to 1600 ms relative to the acoustic onset of the target word. Ocular and muscular artifacts were corrected by principal component analysis (e.g., Wallstrom, Kass, Cohn, & Fox, 2004). Segments with potentials exceeding ± 75 µV were rejected, and the remainder was screened for drift artifacts. Epochs contaminated with artifacts were discarded, leading to an average segment loss of (8.51%) across participants. Trials in which participant responses preceded the response prompt were also rejected (1.1%). Exclusion rates were comparable across conditions. Separate ERP waveforms obtained for each participant in each condition were baseline-corrected using a 200-ms pre-stimulus interval and digitally re-referenced to an average of the left and right mastoid channels.

To determine the time window and scalp distribution of any significant differences between conditions, ERP signals were submitted to repeated-measures, two-tailed cluster-based permutation tests with a family-wise alpha level of 0.05 (Bullmore et al., 1999) using the Mass Univariate ERP Toolbox (Groppe, Urbach, & Kutas, 2011).[3] All time points between 0 and 1600 ms at all 61 non-ocular electrodes were included, yielding 19,520 comparisons in total for each test. To estimate the distribution of the null hypothesis, 2500 random within-participant permutations of the data were used — more than twice the recommended number for an alpha level of 0.05 (Manly, 2006). For each permutation, all *t*-scores corresponding to uncorrected *p*-values of 0.05 or less were formed into clusters with any neighboring *t*-scores meeting the same criteria. Electrodes within 5 cm of one another were considered spatial neighbors and adjacent time points were considered temporal neighbors. The cluster "mass" was defined as the sum of the *t*-scores in each cluster, and the most extreme cluster mass in each of the 2501 sets of tests was recorded and used to estimate the distribution of the null hypothesis. The permutation cluster mass percentile ranking of each cluster from the observed data was used to derive its *p*-value. The *p*-value of the cluster was assigned to each member of the cluster and *t*-scores that were not included in a cluster were assigned a *p*-value of 1. The results of a cluster analysis indicate how many clusters of positive and negative *t*-scores are found for a particular contrast, as well as the range of *p*-values for each. Thus, we report *p*-values for both positive and negative clusters. Note that as when one uses Bonferroni correction, clusters can take a *p*-value of greater than 1.

3. Results[4]

3.1. Behavioral results

To establish whether the presence of a competitor in the privileged or common ground affected behavioral responses as in prior studies (Apperly et al., 2010; Wu & Keysar, 2007), accuracy rates and response times as a function of condition (PGC, CGC, NoC) were analyzed using mixed-effects linear regression with full random effects for participants and items (Baayen, Davidson & Bates, 2008; Barr, Levy, Scheepers, & Tily, 2013).[5] The model for the full data set parameterized the ternary condition variable using orthogonal sum-coded binary predictors. For pairwise analyses, models were fit using a restricted data set containing only the two conditions of interest. Accuracy rates were essentially identical and at ceiling across conditions ($M = 98.7\%$, $SE = 0.15\%$; $ts \ll 1$). However, response times varied by condition such that responses were fastest for the unambiguous NoC ($M = 725$ ms, $SE = 42$ ms), slowest for the referentially-ambiguous CGC ($M = 782$ ms, $SE = 44$ ms), and intermediate for PGC ($M = 752$ ms, $SE = 43$ ms). A model including condition as a predictor fared significantly better than one without ($\chi^2[2] = 7.75$, $p < .05$). Pairwise comparisons indicated that responses to NoC were reliably faster than to CGC [$\beta = 58.5$, $SE = 20$, $t = 2.92$, $p < .01$] and to PGC [$\beta = 28.6$, $SE = 13$, $t = 2.15$, $p < .05$]. Responses to PGC were not reliably faster than to CGC [$\beta = 29.3$, $SE = 20.6$, $t = 1.42$].

This pattern replicates the common finding that listeners are delayed in ultimately selecting the correct target object when a privileged competitor is present. To determine whether this delay is because the privileged competitor is considered to be a referential candidate or is instead due to attentional distraction effects (i.e. the privileged competitor draws non-referential attention away from the target), we turn to ERPs.

3.2. ERP results

3.2.1. CGC vs NoC

Our first aim was to test whether the referentially-ambiguous CGC condition elicited the Nref signature of referential ambiguity. If so, it would establish that the Nref effect is sensitive to referential ambiguity in the present paradigm. We compared ERPs elicited by the CGC condition to the NoC control. This contrast allowed us to directly compare brain responses to the same auditory stimulus while participants viewed visual displays that either made the target in the unfolding utterance temporarily ambiguous (CGC) or unambiguous (NoC).

Fig. 3A shows grand average ERP waveforms time-locked to the acoustic onset of target words. Inherent in the use of natural, fully connected speech stimuli, no exogenous ERP components are evident. However, the CG-NO difference waves (Fig. 3B and C, left) reveal a clear effect of referential ambiguity. CGC elicited a widely distributed negativity from approximately 600–1200 ms, which became progressively more anterior throughout the remainder of the epoch (1200–1600 ms). The scalp distribution of this pattern is consistent with

---

[3] Cluster-based permutation analyses were used in lieu of more conventional mean amplitude ANOVAs because (a) they provide better spatial and temporal resolution than conventional ANOVAs, (b) they have been shown to have good power for detecting broadly distributed ERP effects, and (c) they simultaneously correct for multiple comparisons (Groppe, Urbach & Kutas, 2011; Maris & Oostenveld, 2007). However, we also conducted pairwise repeated measures ANOVAs for all contrasts reported here (conducted separately in three temporal windows: 0–600 ms, 600–1200 ms, 1200–1600 ms), which yielded qualitatively similar results.

[4] Original data from this study are publicly available (Sikos, 2018).

[5] Two items were excluded from behavioral analyses because the correct response could not be determined due to a coding error. To get significance levels, *t*-tests used Satterthwaite approximations for degrees of freedom (Luke, 2017).

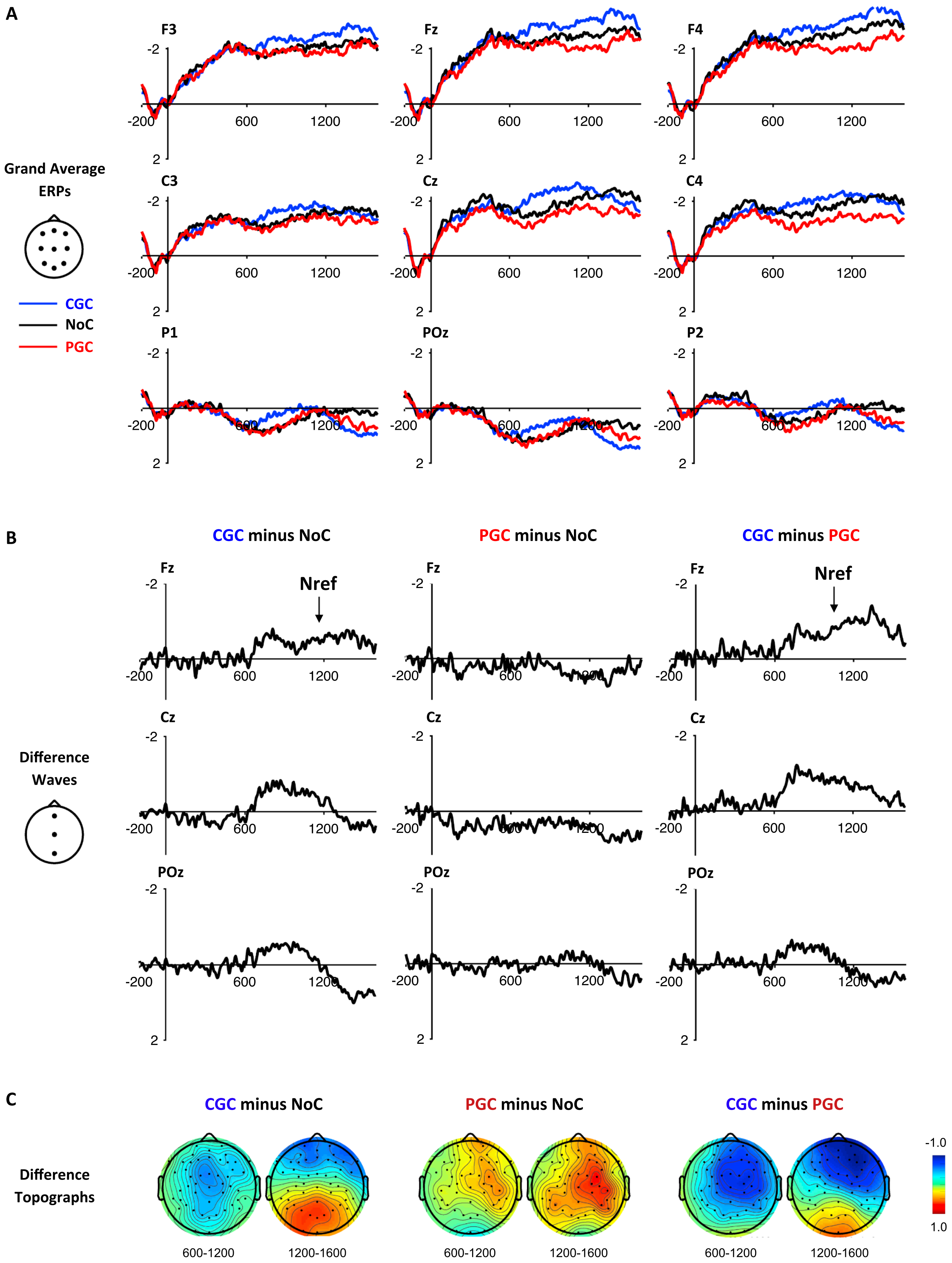

Fig. 3. ERPs. (A) Grand average ERPs elicited by target words in the referentially-ambiguous Common Ground Competitor (CGC) condition, Privileged Ground Competitor (PGC) condition, and unambiguous No Competitor (NoC) condition at nine representative electrodes. Negativity is plotted upwards. (B) Difference waves for the indicated contrasts at three midline electrodes. (C) Difference topographs for the indicated contrasts averaged within the 600–1200 ms and 1200–1600 ms time windows.

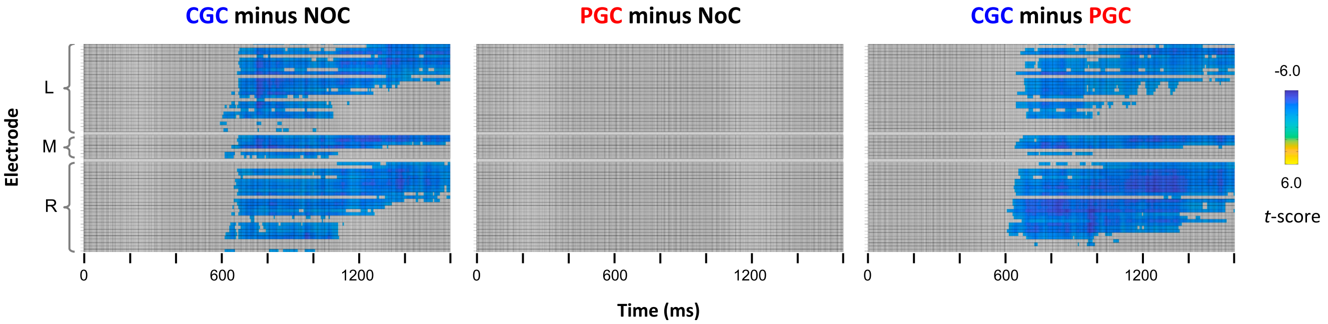

Fig. 4. Statistical results. Raster diagrams visualizing the results of cluster-based permutation analyses for the indicated contrasts. Each cell in a raster diagram represents the result of a t-test. The y-axis indicates electrodes, organized according to laterality and region: electrodes on the left (L) side of the head are grouped at the top, right (R) at the bottom, and midline (M) in the center. Each grouping is ordered top-to-bottom from frontal to posterior electrodes. The x-axis indicates time in 5 ms increments. Color indicates that the test was significant ($p < .05$) after correcting for multiple comparisons. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

previous Nref effects (e.g., Nieuwland et al., 2007; Van Berkum et al., 1999, 2003).

A repeated measures cluster-based permutation analysis directly comparing CGC and NoC revealed one significant negative cluster ($p < .05$) lasting from 595 to 1600 ms and consisting of up to 55 electrodes. This result confirms that CGC elicited a reliable and widely distributed negativity relative to NoC. No significant positive clusters were found ($ps \gg 0.81$). Fig. 4 (left) summarizes the temporal and spatial extent of this effect. Each cell in the raster diagram represents the result of a single t-test. Cells corresponding to tests that reached significance after correcting for multiple comparisons with a test-wise alpha level of 0.0499 are indicated in color.

3.2.2. CGC vs PGC

Having established the spatio-temporal characteristics of the ERP response to referential ambiguity in the present paradigm, we then assessed whether ERPs elicited by the critical PGC condition exhibited a similar pattern by comparing CGC to PGC. The CGC-PGC contrast allowed us to compare ERPs elicited by the same auditory stimulus while participants viewed either a visual display that was uncontroversially ambiguous (CGC), or a visual display with a competitor in privileged ground that might potentially render the target reference ambiguous (PGC).

As depicted in Fig. 3B and C, ERPs elicited by the CGC-PGC contrast are qualitatively identical to the CGC-NoC contrast. A repeated measures cluster-based permutation analysis directly comparing CGC and PGC (Fig. 4, right) revealed one significant negative cluster ($p < .05$) from 605 to 1600 ms and consisting of up to 51 electrodes. No significant positive clusters were found ($ps \gg 1.59$). This result indicates that the common ground competitor elicited significantly greater competition than the privileged competitor and runs counter to the claims of strong egocentricity.

3.2.3. PGC vs NoC

Visually, responses to the PGC condition pattern with the NoC control (Fig. 3B, middle). To assess whether there were any detectable effects of ambiguity in the PGC condition, pairwise repeated measures cluster-based permutation analyses comparing PGC to NoC were conducted as above. No reliable negative clusters ($ps \gg 1.01$) or positive clusters ($ps \gg 0.11$) were found at any point in time nor space (Fig. 4, middle).[6]

3.3. Potential effect of learning

The fact that no differences were found between PGC and NoC stands in stark contrast to the clear and robust differences between CGC and NoC. More importantly, CGC reliably diverges from PGC at the same point in time (605 ms) that CGC diverges from NoC (595 ms). Thus, the results strongly indicate that participants did not consider the privileged competitor to be an equally viable referential candidate as the common ground competitor.

However, because ERP studies generally require a substantially greater number of trials than behavioral studies, one might ask whether the lack of an Nref effect for the PGC condition can be better explained as an effect of learning. That is, could participants have developed a strategy over the course of the experiment to simply ignore competitors in the gray quadrant because they are never the target? This is a potential alternative explanation for the data that might salvage a version of the egocentric account.

There are at least two reasons to believe that this strategic-processing account is unlikely. First, if participants learned to ignore the privileged competitor, then we should have found no difference in mean response times between PGC and NoC conditions. Instead, the behavioral results showed an attentional distraction effect — participants were slower to select the correct target in PGC than NoC trials — indicating that attention was drawn to the privileged competitor and that participants did not learn to ignore the privileged ground.

Second, a quartile analysis of the ERP data revealed that CGC was significantly more negative than both NoC and PGC in the initial quarter of the experiment, and remained numerically more negative for the remainder of the study (Fig. 5A). On a strategic-processing account, the initial pattern of ERP effects should have shown PGC to elicit more negative ERPs than NoC, and little difference from CGC. Moreover, the differences between CGC and PGC should have grown larger across quartiles while the differences between PGC and NoC should have grown smaller. Contrary to these predictions, we found no evidence of egocentricity in the ERP data during any quartile and no qualitative change as the experiment progressed.

For completeness, we also conducted a quartile analysis on the response time data (Fig. 5B). The results of this analysis are consistent with adaptation. The first 50 trials exhibited the same attentional distraction effect as the entire data set, but the effect appears to dissipate over time. Caution should be taken in drawing any firm conclusions from this result without further investigation, however this pattern may suggest adaptation for response times but not for ERPs. This is potential support for the idea that behavioral measures and ERPs can be driven by different cognitive processes. Response time adaptation may be a function of participants becoming more and more familiar with the task over time reducing attentional distraction. There is no indication in this or any other study, that the Nref is affected by such attentional

---

[6] Cluster-based permutation analyses were also conducted comparing CGC to visual control 1 and PGC to visual control 2. These yielded the same patterns as observed for CGC-NoC and PGC-NoC, respectively.

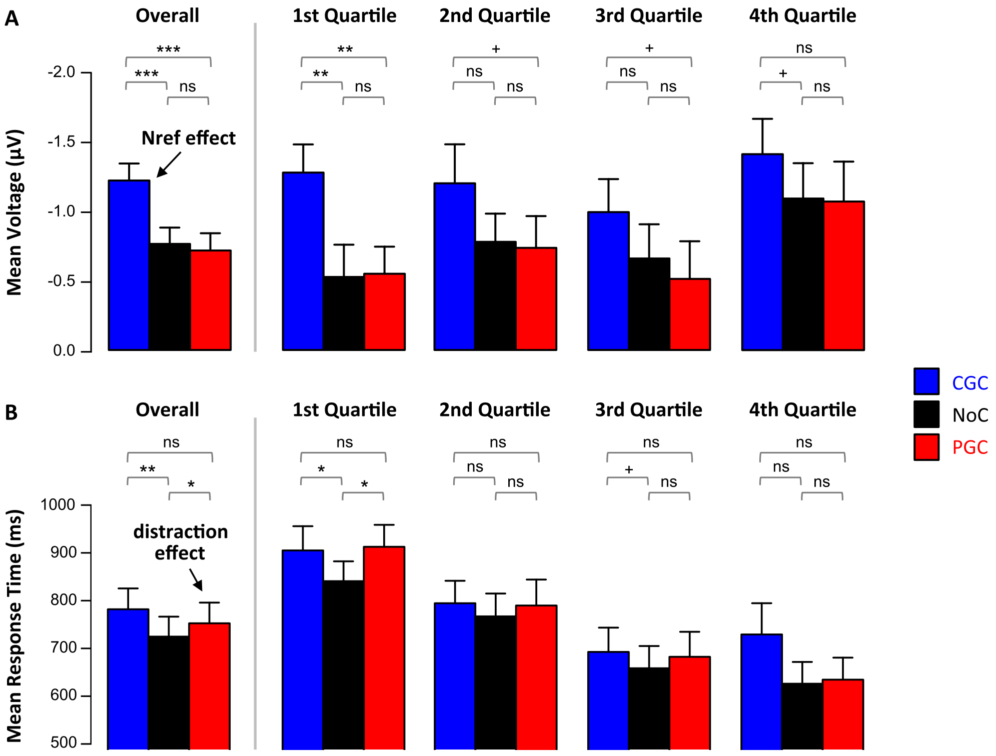

Fig. 5. Quartile analyses. (A) ERP results. Overall mean voltage (μV) by condition (left), and condition by quartile (right), when averaged across all non-ocular electrodes in the 600–1200 ms window. (B) Behavioral results. Overall mean response time (ms) by condition (left), and condition by quartile (right). Error bars represent standard error of the mean. Statistical results from pairwise comparisons are presented above the bar plots. $^{+}p < .1$, $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$.

distraction. Importantly, response time adaptation is common in many experimental paradigms and is quite different from the predictions of the strategic-processing account wherein participants would strategically abandon a "default" egocentric mode of processing by learning to ignore the animal in the gray quadrant.

In sum, both data across the entire experiment and trends in the first quartile of the study militate strongly against the view that participants strategically learned to avoid the referent in the hidden quadrant.

4. Discussion

The present study investigated how quickly ground information can be used to correctly resolve a referent during online processing. Previous studies exploring this question have largely relied on eye-tracking and other behavioral measures. The results of these studies indicate that listeners attend to privileged competitors more than to other privileged objects (Barr, 2008; Keysar et al., 1998, 2000, 2003; Wu & Keysar, 2007). These results have been interpreted as supporting egocentric accounts of referential processing wherein ground information is only taken into consideration during an effortful second stage of processing (Barr & Keysar, 2005; Keysar & Barr, 2002; Keysar et al., 2000, 2003). However, we argue that the behavioral patterns of eye movements and reaction times found in these studies may be confounded with attentional distraction effects. Such effects could arise from phonological and semantic relatedness of the competitor to the target. Behavioral data should therefore be interpreted with caution. To overcome this issue, we measured ERPs and target selection times within a referential communication paradigm. This novel method allowed us to bring electrophysiological evidence to bear on the question.

Behavioral results replicated the attentional distraction effect found in previous studies. Relative to the unambiguous control (NoC), participants' responses were slower when displays contained a privileged competitor (PGC), and were numerically even slower still when displays contained a common ground competitor (CGC). Despite behavioral evidence that participants were affected by both types of competitors, the ERP results showed a dramatically different pattern. The CGC condition elicited robust Nref effects relative to both NoC and PGC. These effects emerged over the same time window and had similar spatial distributions. Moreover, there were no apparent differences between the NoC and PGC conditions. We take this pattern to indicate that while competitors in the common ground were treated as candidates for reference, privileged competitors were not. In other words, participants' expectations about a speaker's referring expressions were quickly constrained by ground information, which precluded a referential ambiguity from arising in the PGC condition.

These findings have important implications for theories of referential processing. The ERP results are incompatible with the Perspective-Adjustment version of the egocentric hypothesis (Keysar et al., 1998, 2000). This account argues that a first-pass egocentric analysis is an automatic and immutable process that cannot be overridden by top-down efforts to ignore objects in privileged ground. However, in the current study we found no electrophysiological evidence for such an early stage of egocentrism. In fact, we found no evidence that participants entertained the privileged competitor as a true candidate for reference at *any point* during the processing of the target word. The Nref effect has a sustained duration and is not extinguished by the resolution of the ambiguity (Nieuwland et al., 2007). Thus even if a period of egocentricity was short lived, it should have been detectable in the ERP record.

On the other hand, the current ERP data could potentially be consistent with the Anticipation-Integration account (Barr, 2008, 2014, 2016) — assuming that a plausible explanation could be identified for why participants overwhelmingly abandoned the "default" egocentric processing mode. For example, Barr (2008) argued that previous eye-tracking studies showing early effects of common ground were the result of such an adaptive strategy because the stimuli that were used in those studies contained fully ambiguous utterances that could only be resolved by considering ground information. Barr (2008) further claimed that "stronger evidence against this account could be obtained by showing that listeners used common ground in a task where the ambiguities were only temporary" (p. 21). The current study provides such evidence because all ambiguities were resolved linguistically before the end of each utterance. Alternatively, one might argue that our participants learned to strategically abandon the default egocentric mode of processing by simply ignoring the privileged ground competitors, and thus attended only to objects in common ground. However, as discussed above, the behavioral results replicate the attentional distraction effect found in previous eye-tracking studies, indicating that participants did not ignore privileged ground objects. Moreover, quartile analyses indicate that the same qualitative pattern of ERP effects was present across the entire study, and was strongest during the initial 50 trials. The strategic-processing account predicts the opposite pattern: that the overall pattern of ERP effects should start weak and become stronger as the study progresses.

The current findings are most consistent with the constraint-based hypothesis. Participants' behavioral responses were slowed by attentional distraction effects when a privileged competitor shared key features with the target, but crucially, the privileged competitor did not elicit the neural signature of referential ambiguity. Thus, listeners appear to encode the status of information as common or privileged, and then use this distinction during online reference resolution to constrain potential referents to only those objects that appear in common ground.

Taken together, the findings reported here demonstrate the value and urgency of using multiple distinct online methods in the study of perspective taking. Previous work in this area has been dominated by eye-tracking in a visual world paradigm. Although eye movements are clearly driven by referential identification processes, they are also influenced by other attentional mechanisms. Our results have demonstrated that behavioral measures and neurophysiological measures can simultaneously show different patterns. Thus it is reasonable to assume that attentional distraction effects — driven by the fact that privileged competitors share key features with the target — will play an important role in this frequently-used paradigm. The novel approach introduced here provides a fruitful method for testing theories of online reference resolution, without being susceptible to such attentional distraction. Moreover, the current findings place important constraints on how cognitive processes may be linked to each type of output measure in referent identification.

Acknowledgments

This research was made possible by a Swarthmore faculty research grant to DJG and by funds from the Swarthmore ERP Lab. We are grateful to Lauren Sanchez and the members of the Swarthmore ERP Lab for their assistance with data collection.

References

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language, 38*(4), 419–439. https://doi.org/10.1006/jmla.1997.2558.

Apperly, I. A., Carroll, D. J., Samson, D., Humphreys, G. W., Qureshi, A., & Moffitt, G. (2010). Why are there limits on theory of mind use? Evidence from adults' ability to follow instructions from an ignorant speaker. *Quarterly Journal of Experimental Psychology, 63*(6), 1201–1217. https://doi.org/10.1080/17470210903281582.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005.

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The autism-spectrum quotient (AQ): Evidence from asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *Journal of Autism and Developmental Disorders, 31*(1), 5–17. https://doi.org/10.1023/A:1005653411471.

Barr, D. J. (2008). Pragmatic expectations and linguistic evidence: Listeners anticipate but do not integrate common ground. *Cognition, 109*(1), 18–40. https://doi.org/10.1016/j.cognition.2008.07.005.

Barr, D. J. (2016). Visual world studies of conversational perspective taking: Similar finings, diverging interpretations. In P. Pyykkönen-Klauck, P. Knoeferle, & M. W. Crocker (Eds.). *Visually situated language comprehension*. Amsterdam: John Benjamins.

Barr, D. J., & Keysar, B. (2005). Making sense of how we make sense: The paradox of egocentrism in language use. In H. L. Colston, & A. N. Katz (Eds.). *Figurative language comprehension: Social and cultural influences* (pp. 21–41). Mahwaw, N.J.: Erlbaum.

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* (pp. 901–938). (2nd ed.). Amsterdam, Netherlands: Elsevier. https://doi.org/10.1016/B978-012369374-7/50024-9.

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. https://doi.org/10.1016/j.jml.2012.11.001.

Barr, D. J. (2014). Perspective taking and its impostors in language use: Four patterns. The Oxford handbook of language and social psychology, 98.

Boudewyn, M. A., Long, D. L., Traxler, M. J., Lesh, T. A., Dave, S., Mangun, G. R., ... Swaab, T. Y. (2015). Sensitivity to referential ambiguity in discourse: The role of attention, working memory, and verbal ability. *Journal of Cognitive Neuroscience, 27*(12), 2309–2323. https://doi.org/10.1162/jocn_a_00837.

Brennan, S. E., & Hanna, J. E. (2009). Partner-specific adaptation in dialog. *Topics in Cognitive Science, 1*(2), 274–291. https://doi.org/10.1111/j.1756-8765.2009.01019.x.

Brodbeck, C., Gwilliams, L., & Pylkkänen, L. (2015). EEG can track the time course of successful reference resolution in small visual worlds. *Frontiers in Psychology, 6*. https://doi.org/10.3389/fpsyg.2015.01787.

Brown-Schmidt (2009). The role of executive function in perspective-taking during online language comprehension. *Psychonomic Bulletin and Review, 16*, 893–900. https://doi.org/10.3758/PBR.16.5.893.

Brown-Schmidt, S., & Hanna, J. E. (2011). Talking in another person's shoes: Incremental perspective-taking in language processing. *Dialogue and Discourse, 2*, 11–33. https://doi.org/10.5087/dad.2011.102.

Brown-Schmidt, S., & Heller, D. (2014). What language processing can tell us about perspective taking: A reply to Bezuidenhout (2013). *Journal of Pragmatics, 60*, 279–284. https://doi.org/10.1016/j.pragma.2013.09.003.

Bullmore, E. T., Suckling, J., Overmeyer, S., Rabe-Hesketh, S., Taylor, E., & Brammer, M. J. (1999). Global, voxel, and cluster tests, by theory and permutation, for a difference between two groups of structural MR images of the brain. *IEEE Transactions on Medical Imaging, 18*(1), 32–42. https://doi.org/10.1109/42.750253.

Clark, H. H. (1996). *Using language.* Cambridge University Presshttps://doi.org/10.1017/cbo9780511620539.

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

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

Dahan, D., Magnuson, J. S., Tanenhaus, M. K., & Hogan, E. M. (2001). Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition. *Language and Cognitive Processes, 16*(5–6), 507–534. https://doi.org/10.1080/01690960143000074.

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods, 134*(1), 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009.

Eberhard, K. M., Spivey-Knowlton, M. J., Sedivy, J. C., & Tanenhaus, M. K. (1995). Eye movements as a window into real-time spoken language comprehension in natural contexts. *Journal of Psycholinguistic Research, 24*(6), 409–436. https://doi.org/10.1007/BF02143160.

Grodner, D., Cheek, N. & Hsieh (2018). Mandatory perspective taking in conversation. Manuscript in preparation.

Grodner, D., Dalini, M., Pearlstein-Levy, S., & Ward, A. (2012). Factors that contribute to the use of perspective in referent identification. In Talk Presented at the 25th annual CUNY conference on human sentence processing.

Groppe, D. M., Urbach, T. P., & Kutas, M. (2011). Mass univariate analysis of event-related brain potentials/fields II: Simulation studies. *Psychophysiology, 48*(12), 1726–1737. https://doi.org/10.1111/j.1469-8986.2011.01272.x.

Hanna, J., & Tanenhaus, M. K. (2004). Pragmatic effects on reference resolution in a collaborative task: Evidence from eye movements. *Cognitive Science, 28*, 105–115. https://doi.org/10.1207/s15516709cog2801_5.

Hanna, J., Tanenhaus, M. K., & Trueswell, J. C. (2003). The effects of common ground and perspective on domains of referential interpretation. *Journal of Memory and Language, 49*, 43–61. https://doi.org/10.1016/S0749-596X(03)00022-6.

Heller, D., Grodner, D., & Tanenhaus, M. K. (2008). The role of perspective in identifying domains of reference. *Cognition, 108*, 831–836. https://doi.org/10.1016/j.cognition.2008.04.008.

Heller, D., Grodner, D., & Tanenhaus, M. K. (2009). The real-time use of information about common ground in restricting domains of reference. In U. Sauerland, & K. Yatsushiro (Eds.). *Semantics and pragmatics: From experiment to theory* (pp. 228–248). Palgrave Studies in Pragmatics, Language & Cognition.

Keysar, B., & Barr, D. J. (2002). Self-anchoring in conversation: Why language users do not do what they should. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.). *Heuristics and biases: The psychology of intuitive judgment* (pp. 150–166). Cambridge, U.K: Cambridge University Press. https://doi.org/10.1017/CBO9780511808098.010.

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. https://doi.org/10.1111/1467-9280.00211.

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. https://doi.org/10.1006/jmla.1998.2563.

Keysar, B., Lin, S., & Barr, D. J. (2003). Limits on theory of mind use in adults. *Cognition, 89*, 25–41. https://doi.org/10.1016/S0010-0277(03)00064-7.

Kronmueller, E., & Barr, D. J. (2007). Perspective-free pragmatics: Broken precedents and the recovery-from-preemption hypothesis. *Journal of Memory and Language, 56*(3), 436–455. https://doi.org/10.1016/j.jml.2006.05.002.

Lopez-Calderon, J., & Luck, S. J. (2014). ERPLAB: An open-source toolbox for the analysis of event-related potentials. *Frontiers in Human Neuroscience, 8*, 213. https://doi.org/10.3389/fnhum.2014.00213.

Luke, S. G. (2017). Evaluating significance in linear mixed-effects models in R. *Behavior Research Methods, 49*(4), 1494–1502. https://doi.org/10.3758/s13428-016-0809-y.

Manly, B. F. (2006). *Randomization, bootstrap and Monte Carlo methods in biology.* Chapman and Hall/CRC.

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods, 164*(1), 177–190. https://doi.org/10.1016/j.jneumeth.2007.03.024.

Nadig, A. S., & Sedivy, J. C. (2002). Evidence of perspective-taking constraints in children's on-line reference resolution. *Psychological Science, 13*(4), 329–336. https://doi.org/10.1111/j.0956-7976.2002.00460.x.

Nieuwland, M. S., Otten, M., & Van Berkum, J. J. A. (2007). Who are you talking about? Tracking discourse-level referential processing with ERPs. *Journal of Cognitive Neuroscience, 19*(2), 228–236. https://doi.org/10.1162/jocn.2007.19.2.228.

Nieuwland, M. S., & Van Berkum, J. J. (2006). Individual differences and contextual bias in pronoun resolution: Evidence from ERPs. *Brain Research, 1118*(1), 155–167. https://doi.org/10.1016/j.brainres.2006.08.022.

Nieuwland, M. S., & Van Berkum, J. J. (2008b). The neurocognition of referential ambiguity in language comprehension. *Language and Linguistics Compass, 2*(4), 603–630. https://doi.org/10.1111/j.1749-818X.2008.00070.x.

Nieuwland, M. S., & Van Berkum, J. J. (2008a). The interplay between semantic and referential aspects of anaphoric noun phrase resolution: Evidence from ERPs. *Brain and Language, 106*(2), 119–131. https://doi.org/10.1016/j.bandl.2008.05.001.

Sikos, L. (2018). Data for: What do you know? ERP evidence for immediate use of common ground during online reference resolution, Mendeley Data, v1. https://doi.org/10.17632/58ftn6t2nx.1.

Van Berkum, J. J. A., Brown, C. M., & Hagoort, P. (1999). Early referential context effects in sentence processing: Evidence from event-related brain potentials. *Journal of Memory and Language, 41*(2), 147–182. https://doi.org/10.1006/jmla.1999.2641.

Van Berkum, J. J. A., Brown, C. M., Hagoort, P., & Zwitserlood, P. (2003). Event-related brain potentials reflect discourse-referential ambiguity in spoken-language comprehension. *Psychophysiology, 40*, 235–248. https://doi.org/10.1111/1469-8986.00025.

Van Berkum, J. J. A., Koornneef, A. W., Otten, M., & Nieuwland, M. S. (2007). Establishing reference in language comprehension: An electrophysiological perspective. *Brain Research, 1146*, 158–171. https://doi.org/10.1016/j.brainres.2006.06.091.

Wallstrom, G. L., Kass, R. E., Miller, A., Cohn, J. F., & Fox, N. A. (2004). Automatic correction of ocular artifacts in the EEG: A comparison of regression-based and component-based methods. *International Journal of Psychophysiology, 53*(2), 105–119. https://doi.org/10.1016/j.ijpsycho.2004.03.007.

Wu, S., & Keysar, B. (2007). The effect of culture on perspective taking. *Psychological Science, 18*, 600–606. https://doi.org/10.1111/j.1467-9280.2007.01946.x.

Yee, E., & Sedivy, J. C. (2006). Eye movements to pictures reveal transient semantic activation during spoken word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*(1), 1. https://doi.org/10.1037/0278-7393.32.1.1.