Do as eye say: Gaze cueing and language in a real-world social interaction

Ross G. Macdonald
University of Dundee, Nethergate, Dundee,
United Kingdom

Benjamin W. Tatler
University of Dundee, Nethergate, Dundee,
United Kingdom

Gaze cues are important in communication. In social interactions gaze cues usually occur with spoken language, yet most previous research has used artificial paradigms without dialogue. The present study investigates the interaction between gaze and language using a real-world paradigm. Each participant followed instructions to build a series of abstract structures out of building blocks, while their eye movements were recorded. The instructor varied the specificity of the instructions (unambiguous or ambiguous) and the presence of gaze cues (present or absent) between participants. Fixations to the blocks were recorded and task performance was measured. The presence of gaze cues led to more accurate performance, more accurate visual selection of the target block and more fixations towards the instructor when ambiguous instructions were given, but not when unambiguous instructions were given. We conclude that people only utilize the gaze cues of others when the cues provide useful information.

Introduction

Eyes are important communicative tools in social interactions. Twelve-month-old infants respond to objects cued by adults' gaze (Thoermer & Sodian, 2001), indicating that eyes have communicative value well before the development of spoken language. After the language system has been mastered, eyes continue to be useful in communication. They can send overt signals with specific messages, such as eye-rolling or winking or can be used more subtly, such as controlling eye contact to show or withhold intimacy (Kleinke, 1986). A key question in studies of the social and communicative role of the eyes has been to consider how gaze cues from one individual influence the allocation of overt and covert attention in another individual: that is, how gaze cues result in gaze following by the observer. In real-world interactions, gaze cues usually occur alongside spoken language. However, most research in this field has used artificial paradigms, usually without dialogue. To understand the communicative value of gaze cueing and gaze following, it is important to consider how gaze interacts with language. This paper investigates this interaction by manipulating gaze and language in a naturalistic setting.

In social settings, in order to utilize and potentially follow gaze cues from others, we must first detect and orient to people's eyes. In images of real-world scenes containing people, observers have a strong tendency to fixate the faces and eyes of people (Birmingham, Bischof, & Kingstone, 2009). When presented with a series of images featuring an actor, and trying to understand the narrative depicted, participants tended to look at the actor's eyes and then to the object that the actor was looking at (Castelhano, Wieth, & Henderson, 2007). This result demonstrates both gaze-seeking and gaze-following behavior in complex, static real-world scenes.

Ricciardelli, Bricolo, Aglioti, and Chelazzi (2002) carried out a key study on gaze following using a Posner-type task (1980). Participants were asked to look to the left or the right, depending on the color of a central instruction cue while their eye movements were recorded. They were asked to ignore an additional central distraction cue, which in this case was a picture of a woman with her eyes manipulated to either look left, right or to the center. The experimenters found that when the distracter cue and central instruction cue were incongruent (that is, the distracting gaze cued the opposite direction to the central instruction cue), participants were more likely to make a saccade to wrong target, following the direction of the distracting

Citation: Macdonald, R. G., & Tatler, B. W. (2013). Do as eye say: Gaze cueing and language in a real-world social interaction. *Journal of Vision*, *13*(4):6, 1-12, http://www.journalofvision.org/content/13/4/6, doi:10.1167/13.4.6.

doi: 10.1167/13.4.6 Received July 27, 2012; published March 11, 2013 ISSN 1534-7362 © 2013 ARVO

gaze cue. This suggests that other people's eyes reflexively guide our own eye direction.

Subsequent authors have refuted this conclusion and found evidence for flexible gaze following behavior. Nummenmaa, Hyona, and Hietanen (2009) found that observers in a virtual-reality walking task moved their eyes in the *opposite* direction to an oncomer's gaze, suggesting that the context of the gaze cue and the observer's goal can affect gaze following. Similarly, Itier, Villate, and Ryan (2007) provide evidence for task-dependent gaze-seeking behavior. Participants were asked to either assess the head direction or eye direction of face stimuli. Although fixations on the eyes were present in both conditions, there were significantly more initial saccades to the eyes in the eye direction condition, suggesting that, rather than automatically orienting towards gaze cues, participants were fixating on the eyes when they were relevant to the task-at-hand. These studies underline the importance of considering the context in which gaze cues are studied. Despite the amount of research investigating whether the attentional mechanisms behind gaze following and seeking are reflexive or flexible (Galfano et al., 2012; Tipples, 2008; Vecera & Rizzo, 2006) there is still no general consensus on this issue. It is likely that both reflexive and flexible processes play a part, but the extent to which each contributes is unclear (Laidlaw, Risko, & Kingstone, 2012; Ricciardelli, Carcagno, Vallar, & Bricolo, 2013). The present study extends the ongoing debate concerning whether gaze seeking and following are more reflexive or flexible in a real world setting.

While the above studies provide crucial evidence regarding how we seek and follow another person's gaze, they consider the influence of gaze cues in the absence of spoken language, thus neglecting a key component of most real-world interactions. Spoken language and gaze allocation are intimately linked: The visual world paradigm (Cooper, 1974) has shown that when people listen to spoken language, they make anticipatory eye movements to objects that relate to the sentence they are hearing (Altmann & Kamide, 1999). Evidence for an effect of gaze cues on participants' understanding of language is provided by Staudte and Crocker (2011): When viewing a video of a robot describing and looking at objects in a scene, utterance comprehension was affected by the gaze behavior of the robot. The authors also found that when correcting statements made by the robot, participants inferred the robot's intention from the direction of its gaze. This result suggests that as well as following gaze cues to relevant objects or areas, people infer meaning from them in a communicative setting.

Knoeferle and Kreysa (2012) demonstrated an interaction between gaze following and language processing. Participants viewed a video that itself contained a screen showing three avatars. During viewing, participants heard sentences describing the avatars on the screen. For each sentence, one avatar was the object and one the subject of the description. Sentences were presented in the common German subject-verb-object (SVO) structure or an uncommon (yet still grammatically correct) object-verb-subject (OVS) structure. Crucially, the video sometimes contained a person watching the screen that displayed the three avatars. This watcher in the video would look from the first referent avatar to the second during the spoken verb. The experimenters found that the participants were quicker to look at the second avatar when gaze cues were present, suggesting a facilitative effect of gaze cueing on sentence comprehension. Interestingly, they found that the facilitative effect of the gaze cues was significantly more pronounced for the common SVO sentences than the unconventional OVS sentences. The authors suggested this result was due to the additional processing difficulty of OVS sentences leaving fewer processing resources to make use of gaze cues. This latter finding suggests that the more difficult language is to process the less listeners will utilize gaze cues. However, the informativeness of the gaze cues here is supportive to the task rather than central to the task: That is, all of the information required to understand the sentence and the relationships displayed was contained in the spoken language. In the present experiment we explore the interaction between language and gaze cues in a situation where gaze cues are informative to the task and required for successful task completion. We consider situations when language does not uniquely specify the object that is required for task completion, but gaze is directed to the required object. Thus identification of the correct object requires use of both the information supplied by language and that supplied by the gaze cues.

While studies using pictures or videos allow experimental control and manipulation of variables in order to understand the interaction between language and gaze cueing, such paradigms may not entirely reflect how gaze cues and language are used in the real-world. Dynamic real-world paradigms show that the extent to which we seek and follow gaze cues may vary depending on how close the paradigm is to a real social interaction. Laidlaw, Foulsham, Kuhn, and Kingstone (2011) found that participants spent less time looking at a present confederate than they spent looking at the same confederate viewed on a monitor. This result was suggested to be due to the potential for interaction with the present confederate. This conclusion has been supported by findings that pedestrians are less likely to look where others walking towards them are looking than where those walking in front of them are looking (Gallup, Chong, & Couzin, 2012). These findings indicate that gaze following may be affected by a desire to avoid the possibility of a real social interaction, which is not taken into account by paradigms with artificial stimuli. When social interactions are actually occurring, Gullberg (2002) found that participants spent less time looking at people's faces and following their gaze when watching a video of a person speaking than when face-to-face with them, suggesting that gaze cues may be of more use in a real interaction. In a more recent study (Freeth, Foulsham, & Kingstone, 2013), participants answered questions posed to them either by an experimenter on a monitor or the same experimenter face-to-face. They found no significant difference across conditions in terms of the amount of time participants fixated on the face of the speaker. However, the experimenters did find that the presence of eye contact resulted in participants looking at the speaker's face for longer in the real-world condition only. This latter finding shows that the effect of a speaker's gaze behavior on the eye movements of a listener is dependent on the speaker being physically present.

The above studies consider gaze cueing in the context of potential and communicative real-world social interactions. Clark and Krych (2004) considered the relationship between language and gaze within the context of an interaction between participants in a collaborative task. In their paradigm, two builders followed the instructions of a director in order to make structures out of Lego blocks. They found that builders would use a wide-range of nonverbal cues to communicate with the director, including gaze cues. These cues led to the directors altering their utterances mid-sentence, responding to the builders' gaze signals. A later study (Hanna & Brennan, 2007) investigated gaze cues in naive instructor/follower pairs during a simple target-matching task. Followers were found to use the gaze cues of directors to identify correct targets before the point of linguistic disambiguation, showing gaze cues to be a useful communicative tool during real one-to-one interactions. By using ecologically valid paradigms these studies have found that gaze cues not only provide meaning in a communicative setting, but also affect the language used in an interaction.

The present study explores how people utilize gaze cues to complete a real-world task as the availability of these cues and the specificity of verbal instructions vary. Similar to Clark and Krych's (2004) paradigm, participants were required to follow instructions to build structures out of blocks. One-to-one interactions between instructor/follower pairs were used, but this experiment differed crucially from Hanna and Brennan (2007) in that the instructor was an experimenter, rather than a participant. This difference allowed for the changes in language specificity and gaze cues to be strictly controlled so we could investigate how participants respond when gaze cues were absent compared to present, as well as observing the effect of altering language specificity on these responses. We varied whether the instructions unambiguously identified a target block or not and simultaneously manipulated whether the instructor supported his spoken instructions with gaze cues. Importantly, participants were not informed about whether the instructor would provide gaze cues and as such we are able to look for spontaneous gaze-seeking and gaze-following behavior. Our paradigm therefore offers a compromise between the ecological validity of studying gaze cueing and language in a real-world interaction, and the experimental manipulation of the informativeness of gaze and language necessary to characterize their interaction in communication.

We considered three aspects of behavior for assessing gaze-cue utilization: gaze seeking, gaze following, and task performance. To assess these, we used eye movement measures that indicate the involvement of each aspect in each instruction (see Methods for details of the measures). For all three measures we were specifically interested in whether the utilization of gaze cues is mediated by the specificity of the information provided in the spoken language. If gaze-cue utilization depends upon the informativeness of the cue, then we might expect greater gaze-cue utilization when the spoken instructions are ambiguous than when they are unambiguous. Flexible cue utilization has been considered in other aspects of eye movement control (Brouwer & Knill, 2007), but it is not known whether the extent to which we utilize gaze cues from another individual in real world social interactions is similarly flexible. This consideration therefore has the potential to reveal new theoretical insights into gaze cueing in real world social interactions.

Methods

Participants

The 16 participants (12 female) were allocated to four experimental conditions (see below). Undergraduate students received course credit for participation.

Materials

Eighty Mega Bloks building blocks were used. They varied in color (23 red, 27 blue, 14 green, and 16 yellow) and shape (11 large [18 mm × 30 mm × 70 mm], 41 small [18 mm × 30 mm × 30 mm] and 18 curved [18 mm × 30 mm × 30 mm]). Blocks were

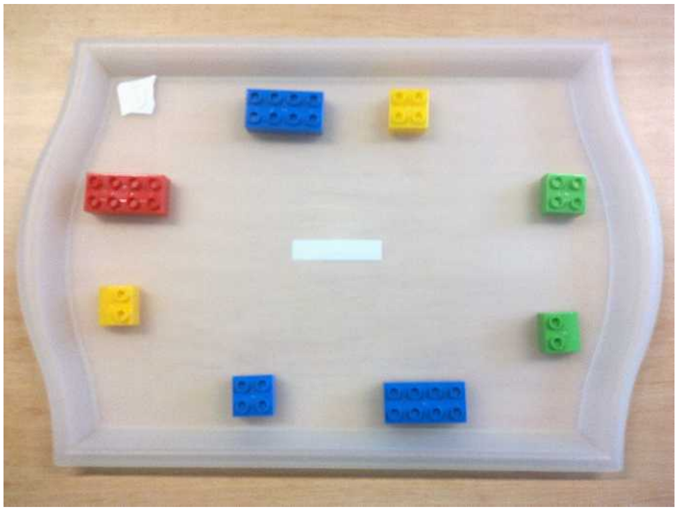

Figure 1. An example of a tray used in the experiment. Four different colors and three different shapes of blocks are positioned in a circle-like arrangement with half to the left and half to the right.

divided across 10 trays (44 cm × 34 cm), with eight blocks per tray arranged in an approximately circular array (Figure 1). The same arrays were used in all conditions, but the order of trays was randomized for each participant.

Design

This experiment had a between-subjects design, with two independent variables: specificity of instruction and gaze-cueing condition. Instructions were either ambiguous or unambiguous. Gaze cues were either provided or not. A between-subjects design was chosen to maximize the opportunity for participants to learn the type of information they will receive from the instructor and so utilize the available information provided by language and gaze. In particular, we wanted to maximize the chance that participants would utilize gaze cues when these were provided.

Procedure

The experimenter and participant sat facing each other throughout the experiment. At the beginning of each set of trials the experimenter placed one of the 10 trays on a desk between himself and the participant. Participants were told to follow the instructions given by the experimenter (referred to as the "instructor") in order to make structures using four of the blocks. The participants were aware that the instructor was an experimenter, but were not made aware that he was investigating gaze cueing. Every set of trials began with the instructor making eye contact with the participant and asking them if they were ready to begin.

Each of the four instruction statements was read in turn by the instructor. The instructor waited until the participant had acted upon the instruction before reading the next one. At the end of each set of trials the instructor provided feedback about whether the model had been completed correctly. Participants were allowed to ask for the instruction statement to be repeated, but could not ask any other questions.

For each instruction there was a clause that specified which block should be picked up, and a clause that specified where the block should be placed. Each pick-up clause and the behavioral response counted as one trial. There were four trials for each structure (set of trials), giving 40 trials overall. The content of the pick-up clause varied between conditions. In the unambiguous instructions condition, the conjunction of features that uniquely defined the target block was provided (e.g., "Pick up the large blue block on the right" for the array in Figure 1). For ambiguous instructions, one of the features of the target object was not given. The unspecified feature was chosen to ensure that the resultant instruction matched two blocks in the array (e.g., "Pick up the large blue block" for the array in Figure 1). In half of the ambiguous pick-up clauses, the color of the block was the feature that was not given and in the other half the location (left of right) was not given. However, while we manipulated the type of information removed, we were not able to consider these two types of information removal separately in our analyses for two reasons. First, the unambiguous sentences varied in whether they contained only featural (shape, size, color) or a mix of featural and spatial descriptions. Second, the spatial cues were always given at the end of the instruction clause whereas featural cues were given toward the start of each clause. Thus comparisons between the two types of removal are not valid. In the gaze-cued conditions, as the instructor read the first word describing a feature of the target block (the first descriptor word), he fixated the target block. This fixation was maintained until the participant picked up a block. This may not be how gaze cues are naturally delivered, but it allowed us to control the quality of the cues between trials, as well as maximize the opportunity for the participants to locate the gaze cues. In the conditions without gaze cues, the instructor looked at his sheet of instructions throughout each trial. Since the participant understood the instructor to be reading from the sheet, this was considered the most natural behavior for the no-gaze condition. Looks to any other point might have been considered to be either socially inappropriate or cues to other locations.

Eye movement recording

Participants' eye movements were tracked using a Positive Science LLC mobile eye tracker (New York, NY), which allowed free head movement. The tracker has two cameras mounted on the frame of a pair of spectacles: one records the scene from the participant's point of view (scene camera), the other records movements of the right eye (eye camera). During the experiment the participant was positioned approximately 110 cm from the experimenter. Data from the cameras was captured in real time on an iMac. Gaze direction was estimated using the Yarbus software provided by Positive Science LLC, which tracks the pupil and corneal reflection. Calibration involved asking the participant to look at certain blocks at certain times. These blocks were positioned on a tray on the table top, about 60–70 cm from the participant's eyes. These calibration blocks were arranged in the same manner as the experimental trays (see Figure 1). After the calibration procedure, tracker accuracy was assessed by asking participants to fixate the blocks again. If the tracker estimates fell within each block, the calibration was deemed adequate; otherwise calibration was repeated. Eye movement data were collected at 30 Hz with a spatial accuracy of about 1°. Sound was also recorded throughout the experiment.

Analysis

Eye-tracking data were manually coded offline by the first author. Saccades were detected manually using deflections of the iris in the video overlay of the eye in the eye tracking video record (for details, see Land & Lee, 1994). The minimum detectable saccade size using this method was 0.5°–1°. There was no minimum fixation duration criterion. The first author coded the timings of all the looks to the blocks and looks to the instructor during the trials. Looks to the instructor were considered to be any fixations that were on any part of the instructor's body. This liberal criterion was used so that we could class any looks to the instructor as potential instances of foveal or parafoveal gaze-seeking behavior. Audacity sound editing software was used to extract the timings for the onset of the first descriptor word. As well as the 16 participants reported here, we tested six other participants, but had to discard these data. Three of the videos had recording errors (e.g., failure to record sound/video) and the remainder had accuracy issues due to calibration errors. Of the 640 experimental trials from the remaining 16 participants, 28 were discarded, due to instruction errors or accuracy problems. We measured three main DVs: (a) percentage of instructions in which the participant looked at the instructor, (b) percentage of trials in which the first block fixated after the onset of the first descriptor word was the target block, and (c) percentage of correct pick-ups of the target block. The percentage of instructions in which the participant looked at the instructor was used as an indicator of gaze-seeking behavior, as participants seeking gaze cues would be expected to look at the speaker more often. Our second DV can be used as a marker for gaze-following behavior. At the point of the onset of the first descriptor word, from language alone, the identity of the target block was ambiguous in all conditions. However, in the conditions with gaze cues, the cues were given at this point, making the identity of the target block unambiguous to any participant following gaze cues. Therefore, the first block to be fixated after the onset of the first descriptor word is more likely to be the target block for those following gaze cues. This measure also has the benefit of taking peripheral gaze following into account: it is possible that gaze cues are followed even when participants did not fixate the instructor, instead detecting the gaze cues using peripheral vision. This measure therefore allows the potential to study aspects of gaze utilization not captured in our gaze-seeking measure. The percentage of correct pick-ups is a measure of task performance that is again able to convey aspects of gaze utilization not available from the other two measures. It may be that gaze cues support behavior and result in more accurate selection of the correct block even in situations where the gaze cues have neither been overtly sought out nor overtly followed by the participant. For all measures, data were analyzed using two (instruction condition) by two (gaze-cueing condition) independent-measures ANOVAs. The time course of gaze following was also investigated by measuring the mean time difference between the first fixation on the correct block and the time of onset of the first descriptor word. This time was selected as it is also the time of onset of the gaze cues in the gaze cue condition. The time course of performance was assessed by measuring the mean time difference between pick-up and the time of the onset of the first descriptor word. Additionally, the temporal evolution of gaze utilization was investigated for those participants provided with gaze cues and ambiguous language. We explored this by measuring the mean time differences between (a) the look to instructor and onset of the gaze cue (b) the onset of the gaze cue and the first fixation on the target block and (c) the first fixation on the target block and the pick-up time.

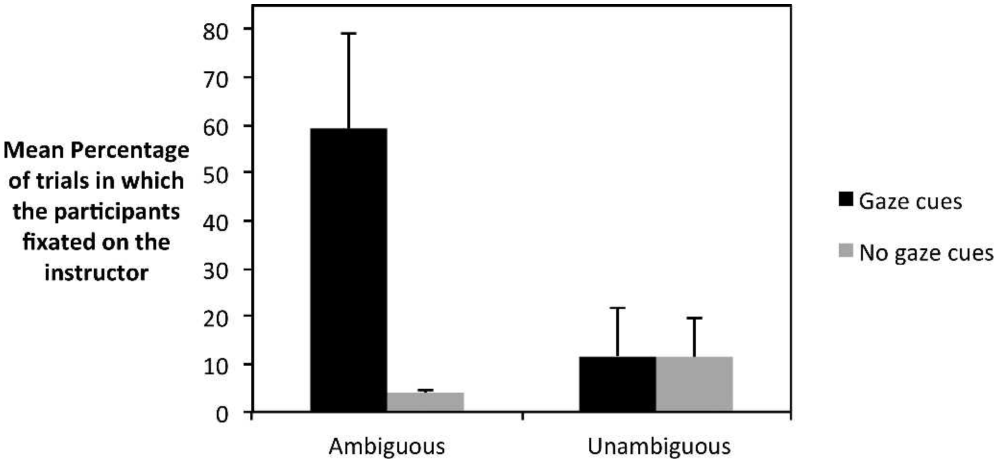

Figure 2. Mean percentage of trials in which the participants fixated the instructor in ambiguous and unambiguous instruction conditions with and without gaze cues (with standard error).

Results

The percentage of instruction statements during which the participant looked at the instructor (Figure 2) was used as an indicator of gaze seeking. There was a main effect of gaze cueing, $F(1,12) = 5.55$, $p = 0.036$, but no main effect of the specificity of language in the spoken instructions, $F(1,12) = 2.90$, $p = 0.114$. Gaze cueing and instruction specificity interacted significantly, $F(1,12) = 5.41$, $p = 0.038$. Planned comparisons showed that for ambiguous spoken instructions, the instructor was fixated more often when he provided gaze cues (59.3% of trials), than when he did not provide gaze cues (4.0%, $p = 0.006$). For unambiguous instructions, gaze cues were rarely sought whether provided by the instructor (11.3%) or not (11.4%); with no significant difference between these conditions ($p = 0.983$).

It is possible that participants in the ambiguous instruction condition may change their gaze-seeking behavior over time, as they gain more experience of the instructor's gaze-cueing behavior. To investigate this, we compared the percentage of trials in which the participants given ambiguous instructions looked at the instructor in the first half and second half of trials. We found no significant difference in first half (56.28%) and second half of trials (62.13%) in the presence of gaze cues, $t(3) = -1.81$, $p = 0.168$. Similarly there was no significant difference between the two halves (4.19%, 3.89%) in the absence of gaze cues, $t(3) = 0.075$, $p = 0.945$.

Gaze following

The first verbal descriptor word for each pick-up instruction never uniquely described the correct block (see Methods). However, gaze cues, which were given at the same time, did specify one block. Any participant following gaze cues would therefore have information specifying the correct block at the point of the onset of the first descriptor word, even though the instruction sentence had not yet reached the point of disambiguation. Given this early advantage, if participants were following gaze cues, it should be more likely that the first block they fixated after the onset of the first descriptor word would be the target block (thus looking to the target block before any other block). Because this measure relies on fixations on the cued location, rather than fixations on or from the gaze cue itself, we are able to consider both overt and covert gaze following behavior, the latter being the detection and utilization of gaze cues without directly fixating on the instructor. Therefore, prior overt gaze seeking is not required to indicate gaze following.

There was a main effect of gaze cueing, $F(1,12) = 6.52$, $p = 0.025$: The first block fixated after the onset of the first descriptor word was more likely to be the target block in the presence of gaze cues than in their absence (Figure 3). There was no main effect of instruction specificity, $F(1,12) = 0.20$, $p = 0.665$. There was a tendency toward an interaction between gaze cueing and instruction specificity, but this failed to reach significance, $F(1,12) = 3.15$, $p = 0.101$.

Since we were a priori interested in whether language mediated gaze-following behavior, we ran planned comparisons despite the lack of significant interaction. There was a significant difference in the mean percentage of trials in which the first block fixated after the descriptor word was the target block in the gaze (64.4%) and no gaze (40.0%) conditions when ambiguous instructions were used ($p = 0.010$), but not when unambiguous instructions were used (gaze cues: 56.9%, no gaze cues: 52.5%, $p = 0.593$). There was no significant difference between the ambiguous and

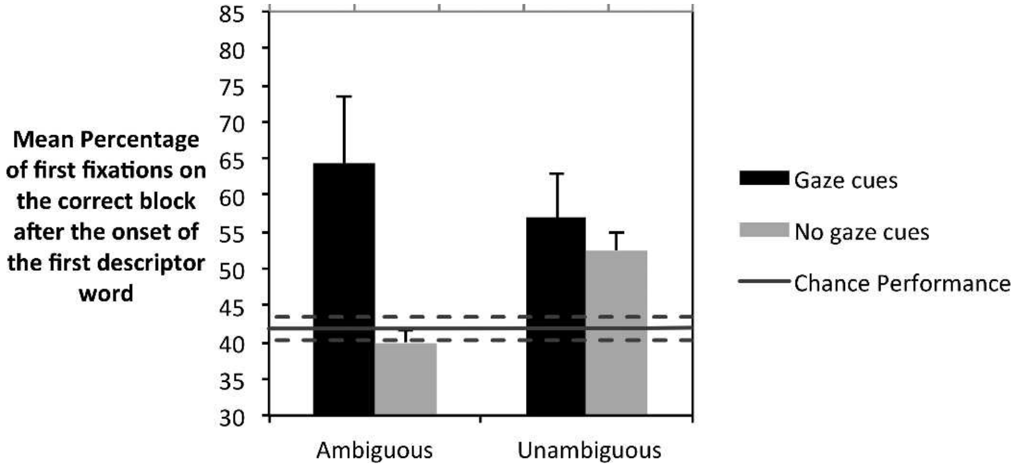

Figure 3. Mean percentage of trials in which the first block fixated after the onset of the first descriptor word was the target block in ambiguous and unambiguous instruction trials with and without gaze cues. The solid horizontal line shows the mean chance of selecting the correct block given only the information supplied by the first descriptor word. Standard errors are shown for all measures. Note that before the onset of the first descriptor word the chance of selecting the correct block would be 12.5%.

unambiguous instructions in the gaze condition ($p = 0.365$), nor was there in the no gaze condition ($p = 0.142$).

The above analyses suggest that in the ambiguous language conditions, gaze cues resulted in an increased likelihood that the target block was the first to be fixated following the first descriptor word (hence following the gaze cue). However, gaze cuing may influence not only the accuracy with which the target block is located, but also the time taken to locate it. To consider this possibility we measured the time difference between the onset of the first descriptor word and the first fixation on the correct block. The variation in instruction length (due to the removal of a descriptor word in the ambiguous conditions) means that these times cannot be compared fairly across instruction condition. However, comparisons were made across gaze condition for each type of instruction. The difference between gaze conditions was not significant with ambiguous instructions (Gaze = 797 ms, $SE = 107$ ms, No gaze = 820 ms, $SE = 32$ ms, $t(6) = -0.21$, $p = 0.840$) or unambiguous instructions (Gaze = 852 ms, $SE = 40$ ms, No gaze = 902 ms, $SE = 22$ ms, $t(6) = -1.10$, $p = 0.316$), suggesting that participants were no quicker to fixate the target block in the presence of gaze cues than in their absence.

Like gaze seeking, it is possible that our gaze-following measure may change over time as participants became accustomed to the instructor's gaze behavior. We investigated this for those in the ambiguous instruction conditions in the same way we investigated gaze seeking. We found that there was a significantly lower percentage of trials in which the first block fixated after the onset of the first descriptor word was the target block in the first half of trials (58.75%) than the second half of trials (70%) for those in the gaze condition, $t(3) = -9.00$, $p = 0.003$. There was no significant difference between the first half (42.5%) and second half of trials (37.5%) in the no-gaze condition, $t(3) = 0.707$, $p = 0.530$.

Task performance

If gaze cues are a communicative tool, then one would expect there to be behavioral benefits of their presence in a task involving communication. The possible benefits of gaze cues on task performance were assessed by comparing the mean percentage of blocks that were correctly picked up (Figure 4). It may be that gaze cues are processed and utilized in ways that are not detected by our first two measures. By measuring task performance in the presence and absence of gaze cues, we are able to see the combined effect of all aspects of gaze utilization. The main effect of gaze cueing approached significance, $F(1,12) = 3.20$, $p = 0.099$, with more blocks correctly selected in the presence of gaze cues. There was a main effect of instruction specificity, $F(1,12) = 52.01$, $p < 0.001$, with more blocks correctly selected when the spoken instructions unambiguously identified the target block. There was a trend toward a significant interaction between these two factors, $F(1,12) = 4.21$, $p = 0.063$. Given our specific theoretical interest in whether gaze cue utilization is sensitive to the information provided by spoken language, we used planned comparisons to address this question. For ambiguous instructions more correct pick-ups were made in the presence of gaze cues (72.1%) than in the absence of gaze cues (52.4%, $p = 0.019$). For unambiguous instructions performance was equally high in the presence of gaze cues (98.7%) and absence of gaze cues (100%, $p = 0.856$).

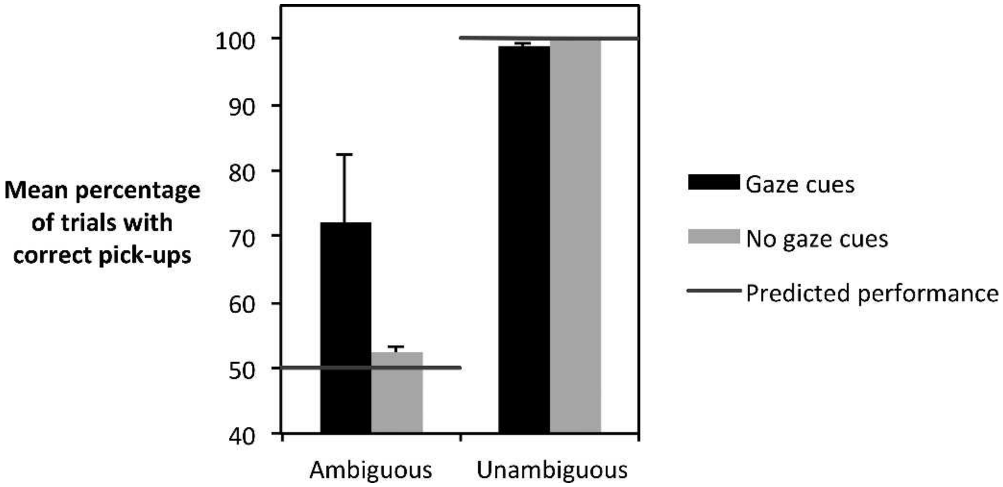

Figure 4. Mean percentage of correct pick-ups (with standard errors) in ambiguous and unambiguous instruction trials with and without gaze cues. The predicted performance from the use of language cues only is shown for both instruction conditions.

Gaze cues may influence not only the accuracy with which blocks are picked up, but also the time taken to pick up the block. To explore this possibility, we measured the time difference between the onset of the first descriptor word and when the correct block was picked up. Independent samples *t*-tests showed that the difference across gaze conditions for this measure was not significant with ambiguous instructions, $t(6) = -0.64$, $p = 0.547$, nor with unambiguous instructions, $t(6) = 0.20$, $p = 0.845$, suggesting that gaze cues did not result in faster times to pick up the correct block.

Temporal evolution of gaze utilization

Each of our dependent variables can be used as a marker for a different aspect of gaze utilization: seeking gaze cues, following gaze cues and task performance benefit for gaze cues. We have argued that each measure has the advantage of capturing different aspects of gaze cue utilization and are not strictly interdependent—for example, gaze following need not require prior overt gaze seeking. However, it is likely that any gaze-seeking behavior should occur before gaze following, which should, in turn, occur before the target block is picked up. It is therefore of interest not only to confirm this temporal ordering, but also to consider the temporal relationships between each stage of gaze-utilizing behavior.

It was only possible to consider the time relations between these three measures for the ambiguous language condition in which gaze cues were present. This is because gaze-seeking behavior was so rare in the other three conditions that no meaningful measures can be taken. Of the 79 trials in this condition that included a look to the instructor, a fixation on the correct block and a correct pick-up, participants looked at the instructor a mean of 896 ms (*SE* 261 ms) before the gaze cue was given. All of the looks to the instructor that occurred before the onset of the first descriptor word in this calculation were maintained at least until the gaze cue was given. After the onset of the gaze cue, participants first fixated on the correct block a mean of 870 ms (*SE* 96 ms) later. Finally, the correct block was fixated a mean of 1136 ms (*SE* 138 ms) before it was picked up. These results show that participants who looked at the instructor would do so before the gaze cue was given; they would then maintain fixation on the instructor until the gaze cue was given and then look to the correct block after the gaze cue was provided. Participants would then typically pick up the correct block roughly one second later. These findings are consistent with the expected viewing pattern of an individual seeking, then following, the gaze cues of the instructor.

Discussion

In a real-world social interaction, participants utilized gaze cues from another individual, but only when the gaze cues provided information that was absent from the spoken instructions. When the information contained in the spoken language was sufficient to uniquely identify a target block, participants rarely sought out or followed gaze cues from the experimenter. However, when language cues were ambiguous, the presence or absence of gaze cues from the instructor did impact behavior in a number of ways: Participants frequently sought out gaze cues when available and they made use of these cues to orient to the target block (gaze following) and to pick up the correct block. Thus our findings add support to the growing body of evidence that gaze-cue utilization is context dependent (Itier et al., 2007; Nummenmaa et al., 2009), and extends this evidence into the domain of a real-world interaction.

The mean percentage of looks to the instructor was significantly higher when the instructions were ambiguous and gaze cues were provided than in any other condition, indicating that these participants were seeking gaze cues more often than all other participants. The difference between this condition and the two unambiguous instruction conditions can be easily accounted for by the irrelevance of gaze cues to the task: The unambiguous language clearly described the location of the target block, so there was no need for any extra nonverbal information. The significant difference across gaze conditions for those given ambiguous instructions is particularly interesting. Those in the condition in which instructions were ambiguous but no gaze cues were provided very rarely looked to the instructor, despite not receiving sufficient verbal information to correctly perform the task. This finding is all the more surprising, considering that the looks to the instructor in the condition in which ambiguous instructions were supported by gaze cues occurred before the gaze cue was given, a point in time at which the gaze and no gaze conditions did not differ. These findings can be explained by the participants very quickly learning the nonverbal informativeness of the instructor: When gaze cues are present participants learned to pre-emptively look to the instructor, however, when they were not present, participants learnt that they cannot receive useful nonverbal information from the instructor and therefore do not look towards him. While these differences must reflect sensitivity on the part of the participant to whether or not gaze cues are provided, we did not see any change in overt gaze-seeking behavior over the course of the experiment, suggesting that participants quickly learned whether gaze cues were provided or were able to detect the presence or absence of gaze cues without overtly orienting to the instructor.

The percentage of trials in which the target was the first block to be fixated after the onset of the first descriptor word was our indicator of gaze following. The identity of the target block was ambiguous at the onset of the first descriptor word to all participants, except those utilizing gaze cues, because the gaze cues were given at this point. We would therefore expect that participants following gaze cues would be more likely to look at the target block before any other block than participants not following gaze cues. Our gaze-seeking indicator can only consider overt gaze seeking, however this indicator for gaze following takes the effects of both overt and peripheral gaze-following into account. This is crucial, as there is evidence that gaze cues can be utilized without being directly fixated (Knoeferle & Kreysa, 2012). Within each gaze condition, there was no difference in the percentage of trials in which the target block was the first block to be fixated after the onset of the first descriptor word for ambiguous and unambiguous instructions. This is likely to be because at the onset of the first descriptor word both types of instruction are equally ambiguous. However, we found a significantly higher percentage of trials in which the target was the first block fixated after the onset of the first descriptor word when ambiguous instructions were supported by gaze cues than when ambiguous instructions were not supported by gaze cues, suggesting that gaze cues were being followed when language was ambiguous. There was no significant difference between the gaze conditions for participants who were provided with unambiguous instructions. We argue that this is due to the gaze cues not being utilized even though they were present, when the necessary information to complete the task was provided by language. Thus, as was the case for our indicator of overt gaze-seeking behavior, gaze following appears to only occur when the language of the instructions was imprecise and did not uniquely identify a target block.

In contrast to our gaze-seeking measure, we found that gaze following appeared to change over the course of the experiment in the condition when gaze cues were present and language was ambiguous. There were more trials in which the target was the first block to be fixated after the onset of the first descriptor word in the final 20 trials than the first 20 trials. This suggests that participants are following gaze cues more when they have encountered more evidence of their communicative value. It is interesting to note that the influence of gaze cues on fixating the target block seemed to be manifest in terms of the accuracy of selecting the target block after the onset of the first descriptor word, but not in terms of fixating this block sooner. There was no difference in how quickly the target block was fixated after the onset of the first descriptor word in the presence or absence of gaze cues. As such, gaze utilization in this experiment appears restricted to selection accuracy rather than selection speed. However, it should be noted that given the small cohort of participants in this experiment, a significant difference in potentially very small time differences would be hard to detect. Taken together, our measures of gaze seeking and gaze following suggest that participants seek and utilize gaze cues when they provide information not present in language, but quickly learn not to seek gaze cues when they are not provided, and do not utilize gaze cues when the spoken instructions contain all of the task-relevant information.

The percentage of trials with correct pick-ups was unsurprisingly very high for both conditions with unambiguous instructions. This was due to the ease of the task when the instructions uniquely identified the target block. Since performance reached ceiling in both conditions, we cannot infer anything from the nonsignificant difference across the gaze conditions. However, we can infer a positive effect of the presence of gaze cues on task performance for participants given ambiguous instructions, as there was a significantly higher percentage of correct pick-ups when ambiguous instructions were supported by gaze cues than when they were not. Although the number of correct pick-ups was higher in the presence of gaze cues, the speed of performance was not significantly affected by gaze cues, suggesting (as for our measure of gaze following) that gaze utilization effects were manifest in terms of accuracy rather than speed of task performance. Our task performance results provide strong evidence that the participants were using the gaze cues to help them in the task.

In all our conditions gaze-seeking behavior was surprisingly rare, with participants only looking at the instructor on around 11% of trials when the spoken instructions unambiguously identified the target block. Even when language was unspecific and gaze cues were therefore highly informative for the participant, they only sought gaze cues on around 59% of trials. This may seem surprisingly low given previous reports of the tendency of people to look at the eyes and faces of others (Birmingham et al., 2009). However, previous research using both real-world and video stimuli has shown that the extent to which we respond to the gaze allocation of others varies with the ecological validity of the paradigm. Specifically, Gullberg (2002) showed increased gaze following when in a real world interaction compared to when viewing a video-taped speaker and Freeth et al. (2013) found that in a real world setting, a speaker's eye contact increased the fixation duration on the speaker's face, but this was not found when participants viewed video stimuli. An explanation for the results of the present study may be provided by work on the effect of the potential for social interaction on the way people look at others (Laidlaw et al., 2011) and the extent to which we follow the gaze of others (Gallup et al., 2012). The present study used a real-world situation in which there was potential for social interaction. This could have led to participants avoiding looking at the instructor, in a similar way to that found by Laidlaw et al. (2011) and Gallup et al. (2012). There are no social consequences of looking at a photograph of a person, so participants in static image experiments (Birmingham et al., 2009) would not show this aversion behavior. Whether or not this accounts for these findings, it is clear that the mere presence of another person is not sufficient to stimulate gaze-seeking behavior, even when they are using gaze to indicate the location of behaviorally important objects.

Hanna and Brennan's (2007) highly naturalistic study found that participants used the gaze cues of instructors to help them understand instructions. In the present study, by controlling the presence of gaze cues and the specificity of instructions, we have found that rather than being a ubiquitous response to a social interaction, the tendency to engage in gaze seeking and gaze following appears to depend upon the informativeness of gaze cues relative to other information. When language provides the necessary information to locate a block, gaze cues are not sought or followed. This suggests that there is an interaction between language and gaze cueing in communication, with gaze becoming more important when spoken language is less effective at communicating a message or idea.

The above conclusion may initially seem to contradict the conclusions of Knoeferle and Kreysa (2012), however there is a clear theoretical distinction between these two conclusions. Knoeferle and Kreysa (2012) found that using a less common (harder) syntactic structure led to a decrease in gaze utilization and concluded that this was due to the extra cognitive resources required to process the less common structure. The present study found that using more ambiguous (harder) instructions led to an increase in gaze utilization. The key difference between these two studies is that the utilization of gaze cues was never necessary to complete Knoeferle and Kreysa's task, whereas it was essential in the present study when ambiguous language was used. When the sentence was harder in Knoeferle and Kreysa's task, the participant would avoid unnecessary supportive information (gaze cues) and focus on the sentence; however, in the present study, the harder instructions require the use of gaze cues in order for the participant to successfully complete the task.

One interpretation of these data is that in the present study, when the language of the instructions was precise, gaze cues do not provide task-crucial information and thus are only task relevant when the instructions were ambiguous. It is clear from previous studies of fixation behavior in natural settings (e.g., Hayhoe, Shrivastava, Mruczek, & Pelz, 2003; Land, Mennie, & Rusted, 1999; see Tatler, Hayhoe, Land, & Ballard, 2011, for a discussion) that fixations are intimately linked to task relevant sources of information in our surroundings. Very few fixations in natural tasks are directed to task-irrelevant objects in the environment. Thus the present data demonstrate that restriction of visual selection to task relevant information extends to social signals conveyed in the eyes of another individual.

The notion that gaze cues are sought and utilized to aid task performance only when they are highly informative is consistent with studies that have considered the extent to which visual cues are used in other domains. When moving virtual objects on a surface, the extent to which hand movements are planned and executed on the basis of visual and remembered information depends upon the relative availability of visual information (Brouwer & Knill, 2007). When walking toward a goal, the relative reliance upon optic flow and egocentric direction depends upon the relative availability of optic flow information (Warren, Kay, Zosh, Duchon, & Sahuc, 2001). It would appear that the extent to which we utilize gaze cues in social interactions may be similarly flexible and dependent upon their informativeness relative to other available cues.

In a similar way to gaze utilization, natural language production can be affected by the relative availability of other information. Brown-Schmidt and Tanenhaus (2008) found that when conversing during a collaborative task, two isolated participants would refer to task-relevant objects with ambiguous descriptive phrases, yet the listener would often have no difficulty understanding what object was being referred to. The authors suggest that the apparently ambiguous statements are produced because the conversational context restricts and aligns the referential domains of the speaker and listener such that for the subset of objects within the shared referential domain, the language is not ambiguous. One interpretation of our findings is that the speaker's gaze can provide similar cues for restricting and aligning referential domains, thus allowing the listener to select the correct referent in spite of the ambiguous language of the instruction. We can therefore speculate that in natural conversation, not only does language affect the utilization of gaze cues, but the informativeness of a gaze cue may reciprocally affect the specificity of language provided.

Our results provide new insights into how gaze cues are utilized in a social setting. In a reasonably ecologically valid social interaction in which a participant follows the instructions of another individual, both language and gaze cues are utilized by the participant to complete the task successfully. However, when language alone can provide all of the information required for successful task performance, gaze cues are not sought out or utilized by the participant. It is, therefore, clearly not the case that gaze seeking and (when possible) following are ubiquitous behaviors in the context of our paradigm. At least in this form of interaction, we appear only to utilize the gaze cues of another individual when they provide information not otherwise available to us.

Keywords: eye movements, joint attention, spoken language, real world, social interaction

Acknowledgments

This work was supported by a Carnegie-Cameron Scholarship and subsequently a joint EPSRC/University of Dundee Studentship, both awarded to Ross Macdonald. We thank two anonymous reviewers for their helpful comments on an earlier version of this manuscript.

Commercial relationships: none.

Corresponding author: Ross G. Macdonald.
Email: rgmacdonald@dundee.ac.uk.
Address: University of Dundee, Nethergate, Dundee, United Kingdom.

References

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

Birmingham, E., Bischof, W. F., & Kingstone, A. (2009). Get real! Resolving the debate about equivalent social stimuli. *Visual Cognition, 17*(6), 904–924.

Brouwer, A., & Knill, D. (2007). The role of memory in visually guided reaching. *Journal of Vision, 7*(5):6, 1–12, http://www.journalofvision.org/content/7/5/6, doi:10.1167/7.5.6. [PubMed] [Article]

Brown-Schmidt, S., & Tanenhaus, M. K. (2008). Real-time investigation of referential domains in unscripted conversation: A targeted language game approach. *Cognitive Science, 32*(4), 643–684.

Castelhano, M. S., Wieth, M., & Henderson, J. M. (2007). I see what you see: Eye movements in real-world scenes are affected by perceived direction of gaze. In L. Paletta & E. Rome (Eds.), *Attention in cognitive systems. Theories and systems from an interdisciplinary viewpoint* (pp. 251–262). Berlin, Germany: Springer-Verlag.

Clark, H. H., & Krych, M. A. (2004). Speaking while monitoring addressees for understanding. *Journal of Memory and Language, 50*(1), 62–81.

Cooper, R. M. (1974). Control of eye fixation by meaning of spoken language: New methodology for real-time investigation of speech perception, memory and language processing. *Cognitive Psychology, 6,* 84–107.

Freeth, M., Foulsham, T., & Kingstone, A. (2013). What affects social attention? Social presence, eye contact and autistic traits. *PLoS One, 8*(1), e53286.

Galfano, G., Dalmaso, M., Marzoli, D., Pavan, G., Coricelli, C., & Castelli, L. (2012). Eye gaze cannot be ignored (but neither can arrows). *Quarterly Journal of Experimental Psychology, 65,* 1895–1910.

Gallup, A. C., Chong, A., & Couzin, I. D. (2012). The directional flow of visual information transfer between pedestrians. *Biology Letters, 8*(4), 520–522.

Gullberg, M. (2002). Eye movements and gestures in human interaction. In J. Hyona, R. Radach, & H. Deubel, (Eds.), *The mind's eye: Cognitive and applied aspects of eye movement research* (pp. 685–703). Amsterdam, The Netherlands: Elsevier.

Hanna, J. E., & Brennan, S. E. (2007). Speakers' eye gaze disambiguates referring expressions early during face-to-face conversation. *Journal of Memory and Language, 57,* 596–615.

Hayhoe, M. M., Shrivastava, A., Mruczek, R., & Pelz, J. B. (2003). Visual memory and motor planning in a natural task. *Journal of Vision, 3*(1):6, 49–63, http://www.journalofvision.org/content/3/1/6, doi: 10.1167/3.1.6. [PubMed] [Article]

Itier, R. J., Villate, C., & Ryan, J. D. (2007). Eyes always attract attention but gaze orienting is task-dependent: Evidence from eye movement monitoring. *Neuropsychologia, 45,* 1019–1028.

Kleinke, C. L. (1986). Gaze and eye contact: A research review. *Psychological Bulletin, 100*(1), 78–100.

Knoeferle, P., & Kreysa, H. (2012). Can speaker gaze modulate syntactic structuring and thematic role assignment during spoken sentence comprehension? *Frontiers in Psychology, 3,* 538.

Laidlaw, K. E. W., Foulsham, T., Kuhn, G., & Kingstone, A. (2011). Potential social interactions are important to social attention. *Proceedings of the National Academy of Sciences, 108,* 5548–5553.

Laidlaw, K. E. W., Risko, E. F., & Kingstone, A. (2012). A new look at social attention: Orienting to the eyes is not (entirely) under volitional control. *Journal of Experimental Psychology: Human Perception & Performance, 38*(5), 1132–1143.

Land, M. F., & Lee, D. N. (1994). Where we look when we steer. *Nature, 369*(6483), 742–744.

Land, M. F., Mennie, N., & Rusted, J. (1999). The roles of vision and eye movements in the control of activities of daily living. *Perception, 28*(11), 1311–1328.

Nummenmaa, L., Hyona, J., & Hietanen, J. K. (2009). I'll walk this way: Eyes reveal the direction of locomotion and make passersby look and go the other way. *Psychological Science, 20*(12), 1454–1458.

Posner, M. I. (1980). Orienting of attention. *The Quarterly Journal of Experimental Psychology, 32,* 3–25.

Ricciardelli, P., Bricolo, E., Aglioti, S. M., & Chelazzi, L. (2002). My eyes want to look where your eyes are looking: Exploring the tendency to imitate another individual's gaze. *Neuroreport, 13*(17), 2259–2264.

Ricciardelli, P., Carcagno, S., Vallar, G., & Bricolo, E. (2013). Is gaze following purely reflexive or goal-directed instead? Revisiting the automaticity of orienting attention by gaze cues. *Experimental Brain Research, 224*(1), 93–106.

Staudte, M., & Crocker, M. W. (2011). Investigating joint attention mechanisms through spoken human-robot interaction. *Cognition, 120,* 268–291.

Tatler, B. W., Hayhoe, M. M., Land, M. F., & Ballard, D. H. (2011). Eye guidance in natural vision: Reinterpreting salience. *Journal of Vision, 11*(5):5, 1–23, http://www.journalofvision.org/content/11/5/5, doi:10.1167/11.5.5. [PubMed] [Article]

Thoermer, C., & Sodian, B. (2001). Preverbal infants' understanding of referential gestures. *First Language, 21,* 245–264.

Tipples, J. (2008). Orienting to counterpredictive gaze and arrow cues. *Perception and Psychophysics, 70*(1), 77–87.

Vecera, S. P., & Rizzo, M. (2006). Eye gaze does not produce reflexive shifts of attention: Evidence from frontal-lobe damage. *Neuropsychologia, 44,* 150–159.

Warren, W. H., Kay, B. A., Zosh, W. D., Duchon, A. P., & Sahuc, S. (2001). Optic flow is used to control human walking. *Nature Neuroscience, 4,* 213–216.