Referent Expressions and Gaze: Reference Type Influences Real-World Gaze Cue Utilization

Ross G. Macdonald and Benjamin W. Tatler
University of Dundee

Gaze cues are used alongside language to communicate. Lab-based studies have shown that people reflexively follow gaze cue stimuli, however it is unclear whether this affect is present in real interactions. Language specificity influences the extent to which we utilize gaze cues in real interactions, but it is unclear whether the type of language used can similarly affect gaze cue utilization. We aimed to (a) investigate whether automatic gaze following effects are present in real-world interactions, and (b) explore how gaze cue utilization varies depending on the form of concurrent language used. Wearing a mobile eye-tracker, participants followed instructions to complete a real-world search task. The instructor varied the determiner used (featural or spatial) and the presence of gaze cues (absent, congruent, or incongruent). Congruent gaze cues were used more when provided alongside featural references. Incongruent gaze cues were initially followed no more than chance. However, unlike participants in the no-gaze condition, participants in the incongruent condition did not benefit from receiving spatial instructions over featural instructions. We suggest that although participants selectively use informative gaze cues and ignore unreliable gaze cues, visual search can nevertheless be disrupted when inherently spatial gaze cues are accompanied by contradictory verbal spatial references.

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

Gaze cues provide useful information during communication. People can accurately and precisely follow gaze cues (Bock, Dicke & Thier, 2008) and can do this from a young age (Hood, Willen, & Driver, 1998). Many lab-based studies have found evidence for an automatic shift in attention in response to gaze cueing stimuli, however there is a growing body of evidence for selective gaze-following behavior, with the context in which a gaze cue is provided affecting the utilization of that cue (Knoeferle & Kreysa, 2012; Liuzza et al., 2011). The most obvious potential influence on gaze following is spoken language; gaze cues are almost always provided alongside language and both language and gaze cues are used to communicate during social interactions. Indeed, gaze cue utilization has been shown to depend upon the informativeness of concurrent spoken language (Macdonald & Tatler, 2013), with gaze cues only being sought out and followed during a real interaction when spoken instructions were ambiguous.

Despite considerable progress over the last few years in our understanding of gaze cues in natural communication and how they are used together with language, at least two key questions remain unanswered.

First, it is unclear whether there is any automatic component of gaze following in real world interactions. While there have been studies suggesting flexible context dependent gaze following (Nummenmaa, Hyönä, & Hietanen, 2009), recent studies have suggested that automatic gaze cueing effects may be restricted to very early responses (Hill et al., 2010; Kuhn & Kingstone, 2009); these effects might not be detected in typical measures used in studies using more complex paradigms, including real world interactions.

Second, typically studies of the interplay between language and gaze cues have considered how our production of or response to language is mediated by gaze information (Hanna & Brennan, 2007), or how the specificity with which language identifies a target mediates gaze cue utilization (Macdonald & Tatler, 2013). Previous studies have not considered whether and how the nature of the information conveyed in speech might influence gaze cue utilization. Specifically, referring expressions can contain featural or spatial determiners that unambiguously specify the target object. Given the inherently spatial nature of a gaze cue, a concurrent featural referring expression offers information additional to and nonoverlapping with that conveyed by a gaze cue, whereas a concurrent spatial referring expression offers information that largely overlaps with the gaze cue.

The present study therefore aims to (a) investigate whether automatic gaze following effects are present in a real-world interaction, and (b) explore how gaze cue utilization varies depending on whether concurrent spoken language refers to objects using spatial or featural disambiguating determiners.

The Automaticity of Gaze Cue Utilization

In a Posner-type (Posner, 1980) task, incongruent static gaze cue stimuli have been shown to impede performance, even if participants know that the gaze cues are not informative (Friesen &
Kingstone, 1998). Further evidence for this automatic attentional
shift has shown that eye movements are also influenced by these
distractor gaze cues (Galfano et al., 2012; Ricciardelli, Bricolo,
Aglioti, & Chelazzi, 2002). Emery (2000) hypothesized that gaze
following was “hard-wired” in the human brain as well as in other
primates. Evidence has since been found that even lower primates (lemurs) exhibit gaze-following behavior (Shepherd & Platt,
2007). The above evidence suggests that reflexive gaze following
has evolved to be innate in humans and other primates.

There is ongoing debate, however, as to whether Posner-type
(Posner, 1980) tasks are an appropriate way to investigate social
attention and gaze cueing, due to the stimuli being far removed
from real world gaze cues (Birmingham, Bischof, & Kingstone,
2009). Whether or not these paradigms are sometimes useful, it is
clear that the context in which a gaze cue is given does have an
effect on how it is utilized. Nummenmaa et al. (2009) carried out
a virtual reality task, in which a virtual figure walked toward the
participant. The participant had to indicate how they would move
in order to avoid colliding with the oncomer. The experimenters
found that, rather than following the gaze of the oncomer, partici-
pants instead looked in the opposite direction. This clearly shows
that the context in which another’s gaze cue is provided can affect
how an observer responds.

It is possible that the seemingly automatic effects found by
Friesen and Kingstone (1998) and the selective effects found by
Nummenmaa et al. (2009) are not contradictory, but rather that they measured different things. The typical reflexive gaze cueing
effect (Friesen & Kingstone, 1998) may occur early and be later
suppressed by strategic top down control (Nummenmaa et al.,
2009). There is previous evidence for the presence of both auto-
matic and flexible gaze-following mechanisms. Kuhn and King-
stone (2009) carried out a gaze cueing paradigm study using
distracting gaze-cueing stimuli that were accurate on 20% of trials;
therefore, these stimuli were counterpredictive, rather than non-
predictive. Incongruent trials resulted in more saccade errors—
thus instances of following the gaze cue—at short stimulus onset
asynchronies (SOAs) (0–100 ms), but the reverse was true at long
SOAs (900 ms). These findings suggest that there is an early
reflexive shift in attention caused by a gaze cue, but within a
second of cue onset, people are able to use top down control to
inhibit gaze following and shift attention in the opposite direction.
Similarly, Hill et al. (2010) found evidence of automatic gaze
following in a Posner-type (Posner, 1980) task when targets were
presented less than 200 ms after gaze cue onset, but selective gaze
following when targets were presented more than 200 ms after
gaze cue onset. As well as gaze following, there is evidence of
volitional and nonvolitional influences on gaze seeking. Laidlaw,
Risko, and Kingstone (2012) presented participants with static
images of faces and instructed participants to not look at either the
mouths or eyes. Compared to free viewing, participants managed
to easily avoid fixating mouths, but struggled with eyes. However,
when carrying out this task on pictures of inverted faces, partici-
pants could equally suppress looks to mouths and eyes. Combined,
these studies suggest orienting to another’s eyes is both selective
and automatic and that the automatic effect may be linked to
holistic face processing. These studies provide good evidence for
the existence of both automatic and selective components of gaze
cue utilization; however, it is still not clear whether the automatic
component has any effect in a real world social interaction, in
which a number of complex factors influence the use of gaze cues.

Finding evidence for a reflexive component of gaze cue utilization
in real world interactions is important from a theoretical stand-
point, in that it would show that theories of gaze following involv-
ing automatic and selective attentional mechanisms are applicable
to real world interactions, rather than simply lab-based paradigms.

### Language and Gaze Utilization

In the real world, spoken language is arguably the main complex
factor that affects the meaning of a gaze cue. Therefore, to under-
stand how gaze cues are used naturally, it is important to consider
the relationship between spoken language and gaze cueing. In a
naturalistic paradigm, two builders followed the instructions of a
director in order to make structures out of Lego blocks (Clark &
Krych, 2004). The builders spontaneously used gaze cues (among
other nonverbal cues) as part of their communication with the
director. Directors tended to alter their utterances midsentence in
response to these cues, showing that gaze cues can affect language
production to speed up communication. Hanna and Breman (2007)
conducted a similar study in which naïve instructor/fol-
lower pairs carried out a simple target-matching task. Followers
used the gaze cues of directors to identify correct targets before the
point of linguistic disambiguation. These results show that the
meaning of a gaze cue is taken into consideration both when
providing and following verbal instructions.

Children as young as three years old have been shown to use the
meaning of a gaze cue to interpret unknown words (Nappa, Wes-
sel, McEldoon, Gleitman, & Trueswell, 2009). The children were
presented with a screen featuring two static cartoon characters on
the bottom and a video of man looking at either of these characters
on the top. As he looked at these characters the man would make
an ambiguous statement including a nonword verb (e.g., “he’s
blinking him”). The children were then asked to define the non-
word. The experimenters found that the children would define
the verb as the action being carried out by the character fixated
by the man in the video. This shows that the children understand
that the gaze cue indicates the intention of the speaker and they use
this understanding to interpret ambiguous language. In an exper-
iment with an adult population, Staudte and Crocker (2011) found
people used the cues of a robot to infer the intentions of this
nonhuman agent. A video showed the robot looking at an array of
objects, while making incorrect statements about them. The par-
ticipants corrected statements made by the robot in a way that
showed they inferred the robot’s intention from the direction of its
gaze. The findings of these two studies show that people use gaze
cues to help interpret spoken language by inferring the intentions
of the speaker.

Not only can gaze cues alter the production or interpretation of
language, but also language can influence the manner in which
gaze cues are utilized. Macdonald and Tatler (2013) used a real
world communicative task, in which language specificity and gaze
cueing were manipulated. Participants followed instructions to
build a series of structures out of building blocks. The instructor
varied the specificity of the instructions (unambiguous or ambigu-
ous) and the presence of gaze cues (present or absent). The
presence of gaze cues led to more accurate performance, more
accurate visual selection of the target block, and more fixations
toward the instructor when ambiguous instructions were given, but
not when unambiguous instructions were given. This study showed that gaze cues were only sought out and followed when they were uniquely useful in the task.

In Macdonald and Tatler (2013), the informativeness of the language was varied by using referring expressions that either unambiguously identified an object or were ambiguous and equally described two objects in the array. In this way, gaze cues provided no additional information to that provided by unambiguous spoken instructions. However, gaze did provide additional and essential information for locating the target object when provided alongside ambiguous spoken instructions. Gaze cues, therefore, either provided entirely redundant information or were the only source of information that reliably signaled the location of the correct object. The relative informativeness of gaze and language in natural interactions will typically vary in more subtle ways than this. In particular, the type of language used can vary the overlap between information conveyed in language and gaze. Gaze cues are inherently spatial cues; they direct attention to the position of an object, but are not capable of providing any information about the features of the object, such as color or shape. Therefore, a gaze cue may be able to communicate the same information as a verbal spatial reference, but not a verbal featural reference, making gaze cues potentially more informative alongside featural references. It is, therefore, important to consider whether people’s gaze seeking and following behavior reflects the increased informativeness of a gaze cue provided with a featural descriptor compared to a spatial descriptor. Characterizing the relationship between language type and gaze cue utilization will offer important insights into not only the manner in which gaze and language are used in communication, but also the degree of flexibility and sensitivity that participants exhibit when faced with gaze cues in real world situations.

The Present Study

In the present study, we employed a real world search-and-retrieve paradigm in which each participant followed instructions to find and locate objects in a room, while wearing a portable eye-tracker. An instructor (the first author) provided instructions to identify objects that the participant had to find and collect in the room. Spoken instructions used either featural or spatial language to identify the target and were provided alongside congruent gaze cues, incongruent gaze cues, or no gaze cues.

The inclusion of an incongruent gaze cue condition was important for considering our first aim: whether there is evidence of a reflexive component to gaze cue utilization in real world interactions. Situations in which participants follow gaze cues that do not reliably signal information about the target (such as gaze cues that cue a direction incongruent with the target) have provided key evidence for an automatic component of gaze following in previous gaze-cueing paradigm studies (Ricciardelli et al., 2002; Kuhn & Benson, 2009). The present study employed two measures that may be particularly suitable for identifying any early effects of incongruent gaze cues that might indicate reflexive overt or covert following. The direction of the first saccade after the onset of a gaze cue allows us to consider whether incongruent gaze cues result in immediate overt gaze following. The time that elapsed between the onset of the gaze cue and the first fixation on the target objects has the potential to reveal any delay in orienting to the target object following gaze cues; any such delay could reveal covert gaze following in the absence of any overt immediate following response revealed by the direction of the first saccade after the gaze cue.

For our second aim of understanding the manner in which the type of information conveyed in language influences gaze cue utilization, we can compare gaze seeking and following when objects are referred to using featural or spatial language. If gaze cue utilization depends upon the unique information conveyed in the gaze cues, we would expect that gaze cues would be sought out and followed more when provided alongside featural language than when provided alongside spatial language. This is because the spatial information conveyed in gaze cues offers a unique information contribution in the case of the former, but not in the case of the latter, accompanying language. Effects of language content may manifest in differences in gaze seeking, gaze following, or both.

Method

Participants

Twenty-four undergraduate students (23 women) from the University of Dundee were recruited for this study. All participants were aged between 18 and 30 years and were European, although data on specific nationality and ethnicity were not recorded. They received course credit for participating. They were split randomly and equally into three groups: one for each of the gaze conditions.

Materials

Forty experimental items and 20 distracter items were used for this experiment. The 40 experimental items were made up of 20 object pairs. The objects in each pair were identical, except for their color (Figure 1). All of the items were used in each of the eight sets of trials (each set involved five trials). For each set of trials one of eight different object layouts was used. Each layout involved the objects positioned equally across four surfaces (two on table tops, two under tables). To counterbalance for any directional biases, half of the participants used an alternative set of layouts, which were the mirror image of the initial eight layouts. The instructor used a unique instruction sheet for each participant. Each trial was counterbalanced for the location, color, and type of determiner used for the target across participants.

Design

This experiment used a 2 (within subjects) × 3 (between subjects) design. The independent variables were type of determiner.

Figure 1. An example of one of the 20 object pairs. These purses differ only in color. See the online article for the color version of this figure.
used (featural or spatial) and gaze condition (absent, congruent, or incongruent). A between-subjects design was used for gaze condition to give participants the opportunity to learn the type of cue to expect from the instructor, so that we could investigate potential gaze seeking strategies. In order for participants to learn, instructor gaze behavior had to be blocked by gaze condition. In a within-subjects design, this would require a large number of trials per participant, making analysis overly laborious. Furthermore, in a within-subject design learned behaviors from one block may influence another. We therefore decided to assign only one gaze condition to each participant.

**Procedure**

In addition to the mobile eye tracker worn by the participant throughout each recording session, two extra cameras were used. One was placed so that it would capture the instructor (experimenter) throughout the procedure and the other was placed so that it would capture the participant. The eye tracker was calibrated (see the Eye Movement Recording section). The participant was positioned in front of a bag facing the instructor, who was approximately four meters away (Figure 2). The participant was aware that the instructor was an experimenter (lead author), but was not made aware that he was investigating gaze cueing. Between the participant and instructor and on either side of them were two tables. There were 15 objects on each tabletop and 15 objects underneath each table, but still clearly visible by both the participant and instructor from their starting positions. Of the 60 objects, 40 were experimental objects and five of these were the targets for the first set of five trials. The counterpart (identical object with different color) to each of these five target objects was positioned on the opposite side to the target. When experimental items were not targets for any given layout, objects were sometimes positioned on the same side as their counterpart (as is the case with the green and yellow Frisbees in Figure 2).

The instructor explained to the participant that they would be told to pick up objects, one at a time. After each instruction they were required to find the object, pick it up, and then put it in the bag, before returning to their starting position. For each instruction the instructor looked at the participant before beginning. Each instruction statement followed the same structure; they began with the words “Pick-up the [object name]” and were followed by either “on the [left or right]” (spatial condition) or “that’s [object color]” (featural condition). After five instructions were given, the participant was escorted behind a partition and the objects were rearranged into a predetermined layout. Eight layouts were used in total for each participant and the order was randomized. Of the 40 instructions, 20 used a spatial determiner (left or right) and 20 used a featural determiner (object color). In the congruent gaze cue condition, the instructor looked at the target object after stating the object name, but before giving the determiner. At the same point in the incongruent condition, the instructor looked at the counterpart to the target object, on the other side of the room. In the no-gaze condition the instructor looked at neither object, but instead looked down to the instruction sheet.

**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). Data from the cameras were recorded on to two camcorders and videos were later rendered using the Yarbus software (Version 2.2.8.1) provided by Positive Science LLC. Gaze direction was estimated using Yarbus, which tracks the pupil and corneal reflection. Calibration involved asking the participant to look at particular points on a pin board at a distance of about three meters, followed by particular toy building blocks (Megabloks) on a counter top at a distance of 70 cm. Calibration was repeated after the task. 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 and a volunteer research assistant. Saccades were detected manually using deflections of the iris in the video overlay of the eye in the eye tracking video (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 looks to the target and counterpart objects as well as the timings of gaze cues. The first author and a research assistant coded the number of looks to the instructor and the length of looks to the instructor for each trial. At the time of coding the research assistant was naïve to the purpose and manipulations of the experiment. The lead experimenter and research assistant initially coded the same movie file and these were compared by the lead experimenter to ensure a consistent and high

---

1 Before the experiment began, the experimenter explained that spatial references would refer to locations with respect to the participant rather than with respect to the experimenter giving the instruction. Thus an instruction to “pick up the purse on the right” would mean the purse to the right of the participant.
quality of coding. 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 each instruction statement, the onset of each object name, and the onset of each determiner word. Of the 960 experimental trials from the 24 participants, 144 were discarded, due to three of the objects being commonly misidentified, leaving 816 potentially usable trials. There was further data loss, which affected different dependent variables in different ways. The number of trials used in the analysis of each dependent variable is noted below.

Our first three dependent variables (DV s) were indicators of gaze-seeking opportunities and focused on looks to the instructor. These were: (a) the proportion of trials in which the participants fixated the instructor at trial on-set (779 analyzed trials), (b) the proportion of trials in which the participants fixated the instructor at the time of gaze cue (744 analyzed trials), and (c) the mean proportion of time in each trial the participant spent fixating the instructor (642 analyzed trials). These DVs were used as indicators of gaze-seeking opportunities because, although looks to the instructor are not necessarily due to gaze seeking, we would expect those that seek gaze cues to look at the instructor more often and for longer. The instructor provided gaze cues after the onset of the object name, but before the onset of the disambiguating word. Gaze cue onset times were identified from the video footage of the instructor (left-most panel in Figure 2). On average, the gaze cue was provided 791 ms after the onset of the object name. Since no gaze cues were given in the no-gaze cue condition, we could not use that as a time point. Instead, we calculated the mean time the gaze cue was given in the incongruent and congruent conditions relative to the onset of the object name and onset of the determiner. We found that, on average, gaze cues were given after .88 of the time between the onset of the determiner word and object name had elapsed. For each trial in the no-gaze condition, we calculated .88 of the time difference between the onset of the determiner word and onset of the object name. We then added this to the time of onset of the object name to calculate the “time of gaze cue.” When calculating the proportion of time spent looking at the instructor, the onset of the instruction sentence was used as the beginning of the trial and the pick-up time was used as the end of the trial.

Our remaining two dependent variables were used as indicators of gaze following. We measured (d) the proportion of trials in which the first saccade launched after the onset of the gaze cue was in the direction (left or right) in which the target was positioned (750 analyzed trials). This allowed us to investigate initial responses to congruent and incongruent gaze cues. The effect of our conditions on performance in the task was investigated by measuring (e) the time difference between the first look at the target and the onset of the trial (705 analyzed trials). Smaller time differences between the first look at the target object and the onset of trial were used as indicators of gaze following, since those who utilized a gaze cue would be expected to fixate the target more quickly than those who did not.

These dependent variables were all analyzed using the lme4 package (Bates, Maechler, Bolker, & Walker, 2014) in the R statistical programming environment (R Core Team, 2013) to run linear mixed effects models (LMMs). For our first two dependent variables we produced LMMs with gaze condition as a fixed factor with three levels. For each of the remaining dependent variables we produced two models, using gaze condition and reference type as fixed factors with two levels. The first model used data from no-gaze and congruent gaze trials only and the second used no-gaze and incongruent trials only. In all models except those for the first two dependent variables, there were always two fixed effects (gaze condition and language reference type), each with two levels. These variables were sum coded in these models. In all LMMs we used the maximal converging random effects structure (Barr, Levy, Scheepers, & Tily, 2013). When these models failed to converge and required simplification of the random effects structure, our approach was to first remove the correlations between random-slopes and intercepts, followed by the intercept terms, starting with the item random effect (if present). If necessary, further simplification was made by selectively removing the slopes of the interaction and main effect terms initially in the item-level random effect structure. Our LMMs for our initial measure used subject only as a random factor, since this measure concerned the initial stage of each trial, before item was identified, therefore meaning that item could not theoretically have any effect on performance. The remainder of our LMMs used subject and item as random factors. When significant or approaching significant interactions were identified in any LMM, post hoc comparisons were carried out by Tukey’s tests using the glht function in the multcomp library (Hothorn, Bretz, & Westfall, 2008).

Results

Gaze Seeking: Looks to the Instructor

Although looks to the instructor do not necessarily imply gaze seeking, increased gaze seeking is likely to lead to increased looking time at the instructor, as well as increased looks at key points in the trial. Our first measure concerns the beginning of each trial. We were interested in any effect the between-subjects gaze conditions had on whether or not participants fixated the instructor at the onset of each trial. The mean proportion of trials in which this occurred for each gaze condition is shown in Figure 3a. Participants were nearly always looking at the instructor when the trial began in both the no-gaze (.97) and congruent gaze (.98) conditions, but fewer trials began with a look to the instructor in the incongruent gaze condition (.76). An LMM using gaze condition as a fixed factor (and subject as a random factor) confirmed that there was no significant difference between the congruent and no-gaze conditions, $\beta = .149, \text{SE} = 1.303, z = .115, p = .909$, but there was a significant difference between the no-gaze and incongruent conditions, $\beta = -2.844, \text{SE} = 1.170, z = -2.431, p = .015$.

The proportion of trials in which the participants fixated the instructor at the onset of the gaze cue was measured (see Figure 3b). In most trials, at the onset of the gaze cue participants in the congruent gaze condition were looking at the instructor (.87). The proportion of looks to the instructor at this point was lower for the no-gaze condition (.67) and lower still for the incongruent condition (.50). An LMM using gaze condition as a fixed factor (and subject and item as random factors) showed the difference between the no-gaze condition and the congruent condition to be
nonsignificant, $\beta = 1.676$, $SE = 1.091$, $z = 1.536$, $p = .124$. The difference between the incongruent and no-gaze conditions was also nonsignificant, $\beta = -0.104$, $SE = 1.067$, $z = -0.951$, $p = .342$, but the difference between incongruent and congruent conditions was significant, $\beta = 2.690$, $SE = 1.076$, $z = 2.499$, $p = .012$.

We measured the mean proportion of trial time spent looking at the instructor for each of our gaze and reference conditions (Figure 4). An LMM for the congruent and no gaze conditions (using subject and item as random factors) showed no significant effect of gaze condition, $\beta = .039$, $SE = .028$, $t = 1.386$, $p = .153$, but a significant effect of reference condition, $\beta = -.017$, $SE = .008$, $t = -2.179$, $p = .029$, and a significant interaction, $\beta = -.032$, $SE = .015$, $t = 2.094$, $p = .037$. Pairwise comparisons showed that the mean proportion of trial time that the participant spent looking at the instructor was significantly greater when instructions used featural language than when instructions used spatial language to refer to the target object in the congruent condition, $p = .011$. However, when no gaze cues were provided by the instructor, the type of language used to specify the target had no influence on how long the instructor was looked at, $p = .999$.

An LMM for the no-gaze and incongruent gaze conditions (using subject and item as random factors) showed no effect of gaze condition, $\beta = -.024$, $SE = .034$, $t = -0.691$, $p = .468$, nor reference condition, $\beta = -0.006$, $SE = .008$, $t = -0.732$, $p = .465$, nor any significant interaction, $\beta = -.010$, $SE = .016$, $t = -0.534$, $p = .483$.

The results from our analysis of looks to the instructor suggest that participants engaged in selective gaze-seeking behavior. Participants looked to the instructor at the onset of the gaze cue more frequently when these cues were anticipated to be congruent and they spent more time looking at the instructor overall when the gaze cue provided the most uniquely informative information (when presented alongside featural language in the congruent gaze condition). Next we investigated whether gaze and reference conditions similarly influenced aspects of gaze following.

**Initial Gaze Following**

We investigated initial gaze following by measuring the mean proportion of trials in which the first saccade launched after the gaze cue was in the direction (left or right) of the target object (Figure 5). The first eye movement was usually in the correct direction in the congruent gaze cue condition for both spatial (.92) and featural (.89) references. There were fewer trials with correct first saccades in the no-gaze/spatial condition (.79) and fewer still in the no gaze/featural condition (.56). An LMM of the congruent and no-gaze conditions (using subject and item as random factors) showed a significant effect of gaze condition, $\beta = 1.530$, $SE = .294$, $z = 5.199$, $p < .001$, and reference condition, $\beta = .686$, $SE = .258$, $z = 2.655$, $p = .008$, but no significant interaction, $\beta = .774$, $SE = .516$, $z = 1.500$.
Despite the lack of significant interaction, we were a priori interested in whether the congruent gaze influenced behavior when language was featural or spatial. Pairwise comparisons showed significantly more initial saccades in the correct direction when spatial language was used compared with when featural language was used in the no-gaze condition (p < .001), but not in the congruent gaze condition (p = .957).

From Figure 5 it appears that there was little difference in the results for the incongruent gaze condition and no gaze condition for both spatial (.74) and featural (.56) references. An LMM of the incongruent and no-gaze conditions (using subject and item as random factors) confirmed that there was no effect of gaze condition, β = −.125, SE = .218, z = −.573, p = .566, nor any interaction between gaze and reference conditions, β = −.249, SE = .397, z = −.626, p = .531. There was, however, a significant effect of reference condition, β = .933, SE = .199, z = 4.700, p < .001, with more initial saccades toward the correct side of the room when the object was specified using spatial language than when it was specified using featural language.

It is possible that performance in this measure may have been mediated by whether or not the participant was fixating the instructor at the point in time in which the gaze cue was given. To investigate this, we repeated our analysis using only the trials in which the instructor was fixated at the onset of the gaze cue. This analysis includes 87% of congruent trials, 67% of no gaze trials and 50% of incongruent trials (see Figure 3b). The analysis of these data replicated the pattern of results found using the full dataset. A similar analysis using only the trials in which the instructor was not fixated at the onset of the gaze cue would not be appropriate, given the small percentage of trials in which this occurred in the congruent (13%) and no-gaze (33%) conditions. However, for the incongruent gaze condition only, we used an LMM to investigate any interaction between reference type and whether the participant looked at the instructor at the onset of the gaze cue. The LMM showed there to be no difference in performance between the trials in which the instructor was fixated and those in which he was not, β = .176, SE = .479, z = .366, p = .714. A significant effect of reference type was found, β = .908, SE = .343, z = 2.650, p = .008, but there was no interaction, β = .445, SE = .661, z = .673, p = .501.

Our analysis of first saccade direction after the gaze cue onset showed that saccades were no more likely to be launched in the wrong direction following an incongruent gaze cue than in the absence of gaze cues (see Figure 5). This is somewhat in contrast to the typically increased proportion of erroneous gaze-following responses to incongruent gaze cues found in Posner-type gaze-cueing paradigms (Ricciardelli et al., 2002; Kuhn & Kingstone, 2009). However, it may be that any distracting effect of the incongruent gaze cue might manifest not as differences in overt responses but as differences in covert attention allocation. Such effects might result in a delayed overt response following an incongruent gaze cue rather than an incorrectly directed overt response. The next analysis explores the possibility of such covert effects of gaze cues by comparing the time taken to first fixate the target, which could have potentially been influenced by attentional effects that did not result in the launching of saccades in the wrong direction.

Gaze Following: Looks to Objects

We investigated the effect of our conditions on performance in this task by measuring the mean time difference between the onset of the instruction sentence and the first fixation on the target object (Figure 6). An LMM of the congruent (spatial = 70 frs, featural = 67 frs) and no-gaze conditions (spatial = 76 frs, featural = 86 frs), with subject and item as random effects showed a significant effect of gaze condition, β = −12.548, SE = 3.781, t = −3.339, p = .003, and a significant interaction, β = 12.410, SE = 6.001, t = 2.068, p = .039. Reference type did not have a significant effect, β = −3.572, SE = 3.000, t = −1.191, p = .211. Using pairwise comparisons, participants were found to take significantly longer to fixate the target in the featural language condition than the spatial language condition when no gaze cues were present (p = .025), however no difference was found between language reference types when congruent cues were present (p = .831).

The LMM of the incongruent (spatial = 84 frs, featural = 83 frs) and no gaze conditions, with subject and item as random effects, showed no significant effect of gaze condition, β = 3.085, SE = 5.297, t = .582, p = .541, nor reference type, β = −4.519, SE = 2.822, t = −1.602, p = .107, but an approaching significant interaction, β = 10.633, SE = 5.861, t = 1.814, p = .074. Pairwise comparisons showed no significant difference between the featural and spatial instructions for those in the incongruent condition, p = .993, and an approaching significant difference for those in the no gaze condition, p = .061.

As with our previous measure, it is possible that the time difference between the onset of the sentence and the first fixation on the target object may have been affected by whether or not the participant was looking at the instructor when the gaze cue was provided. We therefore repeated our analysis for only those trials

\[ p = .134. \]
in which the instructor was fixated at the onset of the gaze cue. The pattern of results in this reanalysis was identical to our initial analysis of the no-gaze and congruent conditions⁷ but differed in two ways for the no-gaze and incongruent conditions: A significant effect of reference was found, $\beta = -8.086, SE = 3.811, t = -2.120, p = .036$, and the interaction between reference type and gaze condition no longer approached significance, $\beta = 10.847, SE = 7.305, t = 1.48, p = .124$. Since we were a priori interested in how verbal reference type may mediate gaze following, we ran planned comparisons despite the lack of a significant interaction. We found no significant difference between the featural and spatial instructions for those in the incongruent condition, $p = .999$, and a significant difference for those in the no-gaze condition, $p = .016$.

For the incongruent condition only, we carried out an LMM on the time difference between the onset of sentence and the first fixation on the target object. The time spent looking at the instructor was used as a covariate. Participants in the incongruent condition took longer to make their first fixation on the target object when they looked at the instructor as he provided the gaze cue (spatial = 88.6 frs, featural = 90.7 frs) compared to when they did not fixate the instructor (spatial = 77.0 frs, featural = 73.3 frs). This difference approached significance, $\beta = 11.351, SE = 5.692, t = 1.994, p = .071$. This finding is limited, however, in what it can tell us about the effect of incongruent gaze cues. The time difference here may simply be a result of the participants who are not looking at the instructor at the onset of the gaze cue, having a head-start on the search, as they begin looking at the tables earlier. This explanation is supported by the results from the no-gaze condition, in which those looking at the instructor at the “onset of gaze” were significantly slower to fixate the target object, $\beta = 16.168, SE = 3.849, t = 4.201, p = .001$.

### Discussion

In the present study we used a search task to explore the effects of the type of verbal reference used in an instruction on gaze cue utilization and to investigate whether there is any automatic component of gaze following in real world interactions. Our paradigm used real gaze cue stimuli and spoken language to direct participants to find and retrieve objects in a real environment and measured indicators of gaze seeking and following. We found clear signs of selective gaze utilization behavior, dependent on the type of information provided by language. However, we also found some indication that gaze cues can affect visual behavior even if the cues are known to be unreliable, suggesting that the nonvolitional attentional effects of perceiving gaze cues can have an effect in a real-world interaction.

### The Effect of Language Type on Gaze Cue Utilization

We found evidence to suggest that the type of language used in an instruction can influence both gaze seeking and following. This influence can be seen when comparing the effects of language type in the absence of accompanying gaze cues to the effects of language type when accompanied by congruent gaze cues.

When language was provided in the absence of gaze cues, participants showed no differences in how long they spent looking at the instructor, but showed clear language effects on both their initial saccade direction and in the time taken to first fixate the target object. Initial responses were more likely to be toward the correct side of the room when spatial language was used to instruct the participant than when instructions employed featural language. Similarly, participants were faster to first fixate the target object when responding to spatial language instructions than when responding to featural language instructions. The language-based difference in the no-gaze cue condition was as expected. When participants in the no-gaze condition were provided with featural verbal references they were no more likely than chance to make the first saccade (relative to a time point chosen to match the typical onset of a gaze cue on the gaze cueing conditions, see the Method section) toward the correct side of the room. However, they performed significantly better when spatial determiners were used. This can be accounted for by some of these initial eye movements occurring after the end of the sentence: in the spatial determiner sentences, each instruction ended with either the word “left” or “right,” indicating the location of the target. Thus, any participants who made their first eye movement after hearing these instructions were expected to move their eyes in the target direction. In the no-gaze condition, participants were slower to fixate the target object when featural determiners were used compared with spatial determiners. This can be explained by the relative ease of the task with spatial references. When the word “left” or “right” was used, half of the objects ceased to be competitors and the search array was halved. No such effect occurred when featural references (“blue,” “red,” etc.) were used.

In contrast to the findings in the no-gaze condition, when instructions were provided alongside congruent gaze cues, the type of language influenced how long the instructor was looked at, but not the initial saccade direction or the time taken to first fixate the target object. The time spent looking at the instructor was used as

---

⁷ Our LMM for congruent and no gaze conditions showed a significant effect of gaze condition, $\beta = -15.326, SE = 2.585, t = -5.929, p < .001$, and an interaction, $\beta = 16.095, SE = 6.054, t = 2.659, p = .009$. The LMM of the incongruent and no gaze conditions showed no significant effect of gaze condition, $\beta = 4.320, SE = 3.872, t = 1.120, p = .253$, nor any interaction, $\beta = 10.847, SE = 7.305, t = 1.48, p = .124$, however the effect of reference type was significant, $\beta = -8.086, SE = 3.811, t = -2.120, p = .036$. 

---

Figure 6. The mean time difference (frames) between the onset of instruction sentence and the first fixation of the target. Results (with standard error) are shown for the no-gaze, congruent gaze, and incongruent gaze conditions for both featural and spatial reference conditions.
an indicator of gaze seeking as participants actively seeking gaze cues were expected to spend longer fixating an instructor providing gaze cues. Participants provided with featural references in the congruent gaze condition spent the most time looking at the instructor, suggesting that gaze cues were sought out the most when featural references were provided alongside an informative gaze cue. This condition was the condition in which the gaze cue was the most useful, relative to the spoken language being used. This is because the gaze cue was a) directed to the correct object, and b) a unique source of spatial information. In the condition where congruent gaze cues were provided with a spatial determiner, the gaze cue still reliably signaled the location of the target, but provided information that was also provided by language, making it less exclusively informative in providing the spatial location of the target object. Our finding that participants looked at the instructor for longer when gaze cues provided information not available from spoken language provides evidence for selective gaze-seeking behavior in real world interactions dependent on the type of information conveyed in language, rather than the specificity (Macdonald & Tatler, 2013) or complexity (Knoeferle & Kreysera, 2012) of language.

Both the initial saccade direction after the onset of the gaze cue and the time taken to fixate the target object were used as indicators of aspects of gaze following. Overall, congruent gaze cues resulted in more initial saccades toward the correct side of the room and faster times to first fixate the target compared to the condition in which no gaze cues were provided, supporting, in the real world, the typical results of a lab-based gaze-cueing paradigm study (Friesen & Kingstone, 1998; Ricciardelli et al., 2002). However, for both of these measures we did not find the featural language advantage that was evident in the no gaze condition. There are at least two possible interpretations of this lack of language type effect in the congruent gaze cue condition. First, it may be that the congruent gaze cues were particularly helpful for trials in which instructions used featural language and thus were utilized more in the featural language trials, essentially offsetting the spatial language advantage found in the no-gaze cue condition. Second, it may be that gaze cues were followed so often when congruent, regardless of referring expression, that language type no longer had an effect on this measure. Whichever reason underlies the lack of language type effect in the congruent gaze cue condition, we did not find the advantage for featural gaze cues that might have been predicted if the summed informativeness of gaze and language cues underpinned participants’ responses. The total information conveyed to the participant was more when gaze cues accompanied featural language (where each cue provided nonoverlapping information to specify the target) than when gaze cues accompanied spatial language (where the information conveyed in each cue was the same).

The above results provide strong evidence for flexible gaze-seeking behavior, with participants looking more at a cue-giving instructor when the cues were more exclusively informative. In both of our gaze-following indicators, performance benefited from spatial referring expressions when no cues were provided. When congruent cues were provided this spatial benefit disappeared. This could potentially be the result of participants following gaze cues more when they were accompanied by featural referring expressions. However, it is also possible that these results arose from the indiscriminate use of congruent gaze cues.

The Automaticity of Gaze Cue Utilization in a Real World Environment

As well as exploring the effect of language type on gaze cue utilization, the present study also investigated whether automatic gaze following effects were present in a real-world interaction. Previous laboratory-based studies have found evidence for automatic gaze following both in terms of an increased probability that an incongruent gaze cue was (erroneously) followed (Ricciardelli et al., 2002) and an increase in the amount of time that participants took to manually respond to the onset of a peripheral target when preceded by an incongruent gaze cue (Friesen & Kingstone, 1998). The former indicates an overt effect, whereas the latter might indicate either overt or covert following of the incongruent gaze cue.

In the present study we found no evidence for overt following of incongruent gaze cues: the mean proportion of trials in which the first saccade launched after the onset of the gaze cue was in the target direction did not differ significantly between those in the no-gaze and incongruent gaze conditions. This, therefore, suggests that our participants did not initially follow anticipated incongruent gaze cues, since the mean proportion of initial eye movements in the correct direction was the same, regardless of whether an incongruent gaze cue or no gaze cue was given. To allow for the possibility that the data in the incongruent condition were skewed by those trials in which the participants did not see the gaze cue, we repeated our analysis using only the trials in which the instructor fixated at the onset of the gaze cue. On reanalysis, the results still did not differ significantly between the no-gaze and incongruent conditions. We also focused specifically on the incongruent condition and investigated whether looking at the instructor at the onset of the gaze cue affected the mean proportion of trials with initial eye movements in the correct direction. We found that participants behaved the same way, regardless of whether or not they looked at the instructor as he provided the gaze cue. These results provide no indication of an automatic overt gaze-following effect. The incongruent gaze cues had no influence on the direction of the first eye movement after the gaze cue was given, showing that these incongruent (and uninformative) gaze cues were ignored. These results support previous findings showing that unhelpful gaze cues are ignored more (Itier, Villate, & Ryan, 2007) and followed less (Macdonald & Tatler, 2013) than those that are helpful.

While our immediate gaze following and time difference measures show similar patterns of behavior in the congruent gaze condition relative to the no-gaze condition, the relationship differs between the no-gaze and incongruent conditions.

The lack of difference between the no-gaze and incongruent gaze cue conditions for initial saccade direction suggests that incongruent gaze cues were having no measurable effect on initial overt orienting. However, there was an indication that there were differences between the no-gaze and incongruent gaze conditions in the time taken to first fixate the target object, which might indicate some degree of (possibly covert) following of incongruent gaze cues. We found an approaching significant interaction between the gaze conditions (no-gaze vs. incongruent gaze) and reference conditions in the mean time difference between the onset of the trial and the first fixation on the target object. An approaching significant difference was found in first fixation time in the
no-gaze condition between the featural and spatial reference conditions, with the featural trials taking longer. As with our immediate gaze following measure, we reanalyzed this dependent variable using only the trials in which the instructor was fixated at the onset of the gaze cue. We found that this difference between reference conditions was now significant in the no-gaze condition. However, in both analyses there was no significant difference between reference conditions for the incongruent gaze condition. This pattern of difference between the no-gaze and incongruent gaze conditions is surprising, given that we found evidence from our immediate gaze-following measure that incongruent gaze cues were ignored. However, these data show that the spatial language advantage seen in the no-gaze condition was absent in the incongruent gaze condition. It is unclear why this should be if the cues were actively ignored in the incongruent condition; however, we speculate that this effect may be due to the two pieces of contradictory spatial information provided to the participants when incongruent gaze cues were used alongside spatial verbal references: in this condition, the verbal spatial reference (“left” or “right”) was accompanied by a gaze cue to the opposite direction. This contradictory gaze cue may be distracting for the participant and inhibit the beneficial effect of spatial language over featural language, despite participants understanding that the cue was not useful. If this is the reason that the spatial language advantage found in the no-gaze condition disappeared in the incongruent gaze condition, then this result may indicate that gaze cues have had a nonvolitional effect on attention in this real-world interaction, even when gaze cues were used selectively. These findings support the conclusions of lab-based studies that found evidence of automatic and selective gaze-seeking (Laidlaw et al., 2012) and gaze-following (Ricciardelli et al., 2013) mechanisms. Our study, however, extends these findings to a real-world environment, in which real gaze cues were used in a real social interaction.

As well as the results from our gaze-following indicators, we also found evidence of strategic gaze-seeking behavior that can provide insight into the automaticity of gaze cue utilization. The proportion of trials in which the participant fixated the instructor at the onset of the instructional sentence showed that participants fixated the instructor at the beginning of most trials. However, there were significantly fewer trials with initial fixations on the instructor in the incongruent gaze condition. This is likely due to these participants avoiding what they anticipate to be unhelpful gaze cues. This shows strategic gaze-seeking behavior; however, it may also suggest an automatic capture of gaze cue stimuli. Participants began each trial standing facing the instructor, so in order to begin a trial by not looking at the instructor, participants would have to look away. If gaze cues were known to be unhelpful and gaze following was a completely selective and strategic process, then there would be no need to avoid these stimuli. Participants also looked at the instructor at the onset of the gaze cue in fewer trials in the incongruent gaze condition compared to the congruent gaze condition. These findings could suggest the participants were looking away because they anticipated a cue that would hinder them in the task. This explanation is in line with the typical findings of the gaze cueing paradigm (Ricciardelli et al., 2002) in which known unhelpful gaze cues disrupt task performance; however, our results clearly showed that participants strategically altered their gaze-seeking behavior due to their expectation of gaze cue reliability.

Conclusion

The present study explored the interaction between gaze cue utilization and the type of referring expression used in spoken instructions and investigated whether automatic gaze following effects were present in a real-world interaction. We found evidence that participants selectively sought out and followed gaze cues, dependent on the type of information conveyed in concurrent language cues, rather than merely the overall informativeness of the content of the language. However, although participants were able to inhibit gaze following when gaze cues were unreliable, we found evidence to suggest that the speed of visual search was nevertheless disrupted when inherently spatial gaze cues were accompanied by contradictory spatial verbal references. This latter finding suggests that the automatic gaze cueing effect found in highly unnatural Posner-type tasks (Ricciardelli et al., 2002; Galfano et al., 2012) could influence gaze utilization in naturalistic real-world interactions. Previous lab-based studies have suggested both volitional and nonvolitional gaze utilization mechanisms and our results now provide evidence for this in a real-world social interaction.

Our results suggest that gaze seeking behavior appears both strategic and very sensitive to the (relative) informativeness of gaze cues. Specifically, we seek out gaze cues when they are known to be informative and we avoid gaze cues when they are known to be unhelpful and gaze cues are oriented to (that is, looked at) for longer when they convey information not present in concurrent spoken language than when they convey the same information as concurrent spoken language. For gaze following, the effects were perhaps less clear. There was an indication that incongruent gaze cues resulted in participants taking longer to first fixate the target object, but only when accompanying (contradictory) spatial language. The spatial language advantage found when no gaze cues were provided was no longer present when concurrent congruent gaze cues were provided; this might indicate that congruent gaze cues provided information that was particularly helpful when spoken language did not convey spatial information.

References


Received June 3, 2014
Revision received October 13, 2014
Accepted November 5, 2014