Eye can't ignore what you're saying: Varying the reliability of gaze and language

Ross G. Macdonald (rgmacdonald@dundee.ac.uk)
School of Psychology, University of Dundee, Nethergate,
Dundee, DD1 4HN, UK

Benjamin W. Tatler (b.w.tatler@activevisionlab.org)
School of Psychology, University of Dundee, Nethergate,
Dundee, DD1 4HN, UK

Abstract

Gaze cues quickly orient attention, but language can affect the extent to which we follow these cues (Macdonald & Tatler, 2013). We investigated how reliability of language and gaze cues affect attention. Participants, provided with gaze and verbal cues, selected one of two potential targets and received immediate feedback. Different combinations of gaze and language reliabilities (50%, 80%, 100%) were used across nine sessions. The most reliable cue available informed participants' decisions. Language was favoured when reliability was equal and cues incongruent. When language cues were 100% reliable, incongruent gaze cues had a larger detrimental effect on performance when they were 80% reliable compared to 50%. When gaze cues were 100% reliable, there was an overall detrimental effect of unreliable language, with performance slower when language was 50% reliable compared to 80%. We conclude that language cues are favoured and cause disruption when unreliable, even when superfluous to the task.

Keywords: social attention; eye movements; joint attention; language; gaze cueing.

Introduction

During an everyday social interaction, we provide verbal and non-verbal cues to express our intentions, thoughts and emotions. In the research areas of social cognition and attention, cues provided by the gaze of another have been the focus of much empirical research. Gaze cues have been shown to orient attention and viewer gaze automatically, leading some to conclude that humans have evolved to reflexively orient attention in the direction of a gaze cue (Emery, 2000). However, there is evidence that the effects of gaze cues on attention and eye movements can be modulated by the perceived reliability of the cue (Hill et al., 2005). Furthermore, the verbal cues that often accompany gaze cues in the real world have been shown to affect the extent to which people seek and follow the gaze of another (Macdonald & Tatler, 2013). The present study aims to bring together these two factors to investigate how varying the reliability of both gaze cues *and* language cues interact to affect attention and behaviour.

Friesen and Kingstone (1998) developed a Posner-type (1980) gaze-cueing paradigm to investigate the effect of distracting gaze cues on attention. The paradigm involved detecting, localising and identifying a target (a letter – "F" or "T") on the left or right of a display screen. In the centre of the screen there was a simple drawing of a face looking left, right or centrally, which participants were informed did not predict target location. Participants were slower to detect, localise and identify the target when the cue was directed to the invalid location. A later study (Ricciardelli, Bricolo, Aglioti, & Chelazzi, 2002) found that these invalid gaze cues also lead to eye movements in the incorrect direction. The above experiments provide strong evidence for a reflexive shift in attention caused by gaze cues.

The automatic capture of attention, however, cannot fully account for the way we respond to gaze cue stimuli, as these responses have been shown to vary with perceived cue-reliability. When gaze cue validity is reduced to 20%, the reflexive effect is still present, but only at short SOAs (~100ms) (Kuhn & Kingstone, 2009). Hill et al. (2005) directly compared responses to different gaze cue reliabilities. Half of their participants took part in a version of the gaze cueing paradigm in which all gaze cues were either valid or invalid. The remaining participants took part in a task in which 50% of gaze cues were valid and the rest invalid. The invalid trials had a detrimental effect on response time for participants in the former task at SOAs up to 150ms only, however in the latter paradigm the detrimental effect was apparent up to 750ms. The authors argue that this is evidence for two streams of attentional control when viewing a gaze cue. Initially there is an automatic orienting effect (present in both tasks) and then a slower, top-down, selective effect (present when cues were 100% invalid). Whether both, one or neither of these attentional effects is unique to social cues is unclear (for discussion see Birmingham & Kingstone, 2009), however it is clear that the perceived reliability of a gaze cue modulates our response.

Language has also been shown to modulate the utilisation of gaze cues. Linguistic research using the visual-world paradigm has shown an intimate link between the words people hear and where they look; people will not only look to areas referenced in language, but also make anticipatory looks to items relating to sentences they are hearing (Altmann & Kamide, 1999). Research looking specifically at the use of gaze cues in a real-world interaction (Hanna & Brennan, 2007) found that gaze cues facilitate communication; listeners in a target selection task used gaze cues to identify targets before linguistic disambiguation. In a more controlled experiment, Staudte and Crocker (2011) showed participants videos of a robot providing gaze cues to visible items while they heard incorrect sentences about these items. The sentences could always be corrected in two different and equally plausible ways. Participants mostly corrected the sentences in the way that made the gazed-at item the object of the sentence. These studies clearly show

that gaze cues can influence the understanding of ambiguous language.

As well as gaze cues influencing language, the type of language used can influence the utilisation of gaze cues. Knoeferle and Kreysa (2012) found that when unambiguous sentences were harder to process (sentences with an uncommon, but grammatically legal structure), participants were less likely to use helpful gaze cues. Knoeferle and Kreysa argue that the extra cognitive resources required for sentence comprehension leave fewer resources free to utilise non-verbal cues, suggesting a hierarchical use of language and gaze, in which linguistic processing takes precedent over gaze following.

In a previous study, we investigated the interaction between gaze cues and language in a controlled real-world paradigm (Macdonald & Tatler, 2013). Participants followed instructions provided by an experimenter in order to build abstract structures out of building blocks. The experimenter varied the language used (unambiguous or ambiguous) and the presence of gaze cues (present or absent) between participants. Participants were found to only seek and follow gaze cues when language was ambiguous, which was the only condition in which gaze cues were required for the task. We argue that this is strong evidence for selective utilisation of gaze cues, dependent on the informativeness of language.

The above studies show that language has a clear influence on gaze following as does the perceived reliability of a gaze cue. In the present study, we investigate the relationship between gaze and language reliabilities. Participants each carried out 4,320 trials (over nine sessions) of a simple target selection task. We varied the reliabilities of the gaze and language cues and analysed how these cues influenced eye movements and performance.

Given the well-established gaze cueing effects using similar paradigms (Friesen & Kingstone, 1998; Ricciardelli et al, 2002), we hypothesised that gaze cues would influence eye movements even when less reliable than language. However we also hypothesised that this effect would be modulated by gaze cue reliability, as top-down social-attentional processes will inhibit gaze following at low reliabilities (Hill et al, 2005). Previous evidence suggests that gaze cues are ignored in favour of language when gaze provides no additional information (Macdonald & Tatler, 2013) or when available cognitive resources are focussed on language processing (Knoeferle & Kreysa, 2012). Therefore, language cues were predicted to be favoured over gaze cues and lead to task disruption when incongruent.

Methods

Participants

Five students from the University of Dundee took part in nine one-hour sessions over a two-week period. They each received £20 for their participation.

Materials and eye movement recording

Gaze cue stimuli were composed of eight four-second videos. The videos began with a face staring forward, then after two seconds the face turned to look left or right. Each of these clips was mirrored to provide 16 unique clips (eight cueing to the left and eight to the right). Language cues were made up of 30 descriptor audio clips ("select the [object name]"), two spatial-determiner audio clips ("on the [left or right]") and twelve featural-determiner audio clips ("that's [colour]"). On either side of the head stimuli there was a target picture. These pictures were 200x200 pixels in size and featured objects that differed only in colour. There were 30 pairs of pictures, each of which was used once in each 30-trial block. Stimuli were presented on a 19-inch computer monitor (with a resolution of 1024 x 768), approximately 64 cm away from the participant. At this distance the screen was 31.8° x 23.8°of visual angle. A control pad with a left and a right trigger was used for participant responses. This experiment was carried out using an SR Research EyeLink 1000 desk-mounted eye-tracker and the SR Research Experiment Builder software. This system uses corneal reflection and pupil position to calculate where a participant is fixating. Calibration involved the participant fixating on nine markers on the screen. Once calibrated, a verification procedure took place. Verification (and if necessary, re-calibration) was carried out after every block of 30 trials. The mean calibration error was .369° (SD = .260°) of visual angle. Before each trial, participants fixated a marker in the middle of the screen. The lead experimenter could see the estimated fixation point on their display and was required to accept this fixation in order for the trial to begin. The average error for the experiment was .502° (SD = .260°) of visual angle

Design

The present experiment used a within-subjects design. Between sessions there were two independent variables: The probability that the gaze cue was correct (50%, 80%, 100%) and the probability that the language cue was correct (50%, 80%, 100%). Each of the nine sessions used a unique combination of these probabilities. Within each session, there were 16 blocks of trials and in each of these the proportion of correct cues matched the proportion for the whole session. Within each block we varied whether each cue was correct (except when the cue was 100%).

Figure 1: Stills from an example trial. After the central fixation point was fixated, a video began showing a face looking at the participant, with pictures on either side that differed in colour. While this video played the first part of the sentence was heard. After two seconds the head began to turn towards one of the targets while the second part of the sentence was heard. After making a decision the participant received immediate feedback

Procedure

Each participant was initially informed that they would be required to take part in nine sessions over a two-week period. The lead experimenter explained to the participant that in each session they would have to do the same thing; complete 480 trials of a simple decision task. Participants were informed that in each trial they would be shown pictures of two objects on a screen, one on the left and one on the right, and that they would have to choose between these by looking at their chosen target and then pressing the left or right trigger on the control pad. They were told that there would be verbal and non-verbal cues to help guide them, but that these might not always be reliable. Participants were then informed that they would receive immediate visual feedback after each response and that they should try their best to get as many correct as possible. The eye-tracker was then set-up, calibration carried out, and then the first trial began. The trial started with one of the 16 four-second videos playing in the centre of the screen. Throughout the trial the target items were displayed at either side of the video. Alongside the video, one of the two-second descriptor audio clips played. When the audio clip finished (simultaneously with the onset of the gaze cue), the audio determiner clip began playing (Figure 1). After this point, participants were able to select a target, after which they were informed (on screen) whether they were "correct" (in green) or "wrong" (in red). Participants were free to take breaks at any time. The session was complete after 480 trials. Participants returned on eight more occasions to repeat this procedure. Across the sessions language cue reliability (50%, 80%, 100%) and gaze cue reliability (50%, 80%, 100%) varied in all possible combinations. Participants were not told the reliabilities of the gaze and language cues.

Analysis

We used three dependent variables in our analysis: two performance measures (accuracy and response time) and one eye movement measure (time to first fixate the target). Our initial analysis focused on how these variables were affected by the overall reliability of both language and gaze cues. To analyse this we performed 3 (gaze reliability) X 3 (language reliability) ANOVAs for each of our dependent variables.

We also investigated the interaction between cue congruity and reliability. To do this we analysed the session with 80% reliable gaze and language cues. We used this session because it was the only session with incongruent trials in which both cues were a) more informative than chance and b) equally reliable. Secondly, we analysed sessions in which one cue was 100% reliable and the other cue less reliable to see if there was a detrimental effect of incongruent gaze and language cues and to investigate whether this was mediated by cue reliability. For many of these analyses we compared uneven sample sizes, so we avoided using traditional ANOVA models. Instead, we used the *lme4* package in the R statistical programming environment to run linear mixed effects (LME) models, using subject and item as random factors. We carried out model comparisons to calculate *p*-values for main effects and interactions. Post-hoc comparisons were carried out by Tukey tests using the *glht()* function in the *multcomp* library.

Results

Effect of changing cue reliability

Each of the nine sessions used a different combination of language and gaze cue reliabilities. Figure 2 shows the results for our three dependent variables for each session. A clear interaction can be seen between language and gaze cue reliability for accuracy (Figure 2a). A 3x3 ANOVA confirmed main effects of language reliability, $F(2,36)$ =

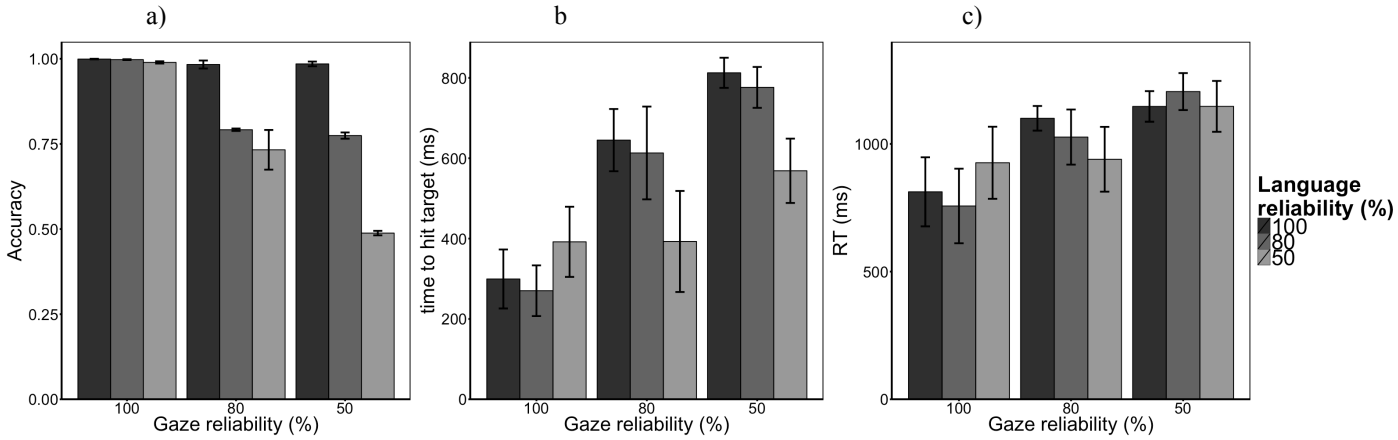

Figure 2: The mean values for a) Accuracy, b) Time to first fixate the target (FT) and c) Response time (RT) across gaze and language reliabilities. Error bars show standard error of means.

115.32, $p < 0.001$, and gaze reliability, $F(2,36) = 112.76$, $p < 0.001$, as well as a significant interaction, $F(4,36) = 38.00$, $p < 0.001$. Accuracy approached ceiling whenever at least one cue was 100%, approached .8 when the most reliable cue was 80% accurate and approached .5 when both cues were 50% accurate. One participant who behaved atypically can account for the large error bar in the 80% gaze cue/ 50% language cue reliabilities condition.

The time to first fixate the target (FT) (Figure 2b) increased as gaze reliability decreased, $F(2,36) = 17.28$, $p < 0.001$. In Figure 2b, FT appears quickest for low reliability language conditions, except when gaze was 100% accurate. However, neither a significant interaction, $F(4,36) = 1.83$, $p = 0.145$, nor any main effect of language reliability was found, $F(2,36) = 2.13$, $p = 0.134$.

Response time similarly increased as gaze reliability decreased, $F(2,36) = 7.01$, $p = 0.003$ (Figure 2c). There was no main effect of language reliability, $F(2,36) = 0.03$, $p = 0.965$, nor any interaction, $F(4,36) = 0.61$, $p = 0.661$.

Effects of cue congruity

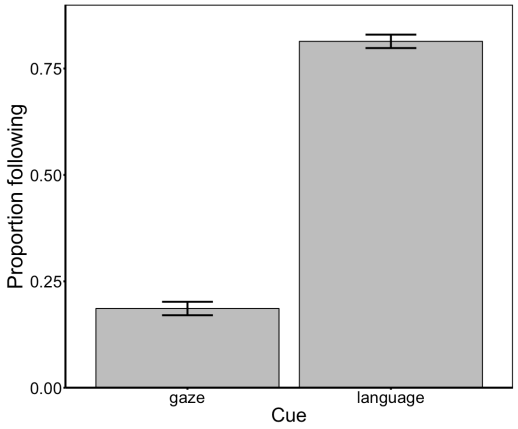

Figure 3: The proportion of cues that were followed in 80% gaze and language cue reliability condition when cues were incongruent. Standard error across trials is shown.

Effects of equally reliable incongruent cues We analysed the session with 80% gaze and language reliability. In this condition, it was equally advantageous to follow gaze or language cues when there was an incongruity. This analysis allowed us to identify any biases for either cue. One participant was removed from this analysis as they exhibited very different behaviour to the other participants. Figure 3 shows the proportion of trials in which each type of cue was followed for incongruent trials only. Language cues were followed in significantly more incongruent trials, $t(6) = 8.627$, $p < 0.001$. This indicates that when cues were incongruent, participants were more likely to follow the language cue.

Effects of the less reliable cue The next stage of our congruity analysis focused on the four sessions in which one cue was 100% accurate and the other cue less accurate. We were specifically interested in whether the reliability of the latter cue affected our measures, even though the former cue was completely reliable. In these four sessions, there were 2-levels of congruity: 1) cues congruent with each other and correct and 2) cues incongruent with each other, with the less reliable cue incorrect.

Figures 4a and 4b show the time to first fixate the target (FT) for congruent and incongruent trials in the four sessions. An LME model of FT (using congruity and gaze reliability as fixed factors) for trials where language reliability was 100% (Figure 4a) showed congruent trials to have significantly quicker FT scores, $\beta = -277.36$, $SE = 39.22$, $t = -7.071$, $p < 001$, as well as an approaching significant interaction between congruity and gaze reliability, $\beta = 234.29$, $SE = 112.74$, $t = -2.078$, $p = .056$, although there was no overall effect of gaze reliability, $\beta = 83.52$, $SE = 68.18$, $t = -1.225$, $p = .207$. Post-hoc Tukey tests showed significant differences in FT between the gaze cue reliability conditions for both congruent and incongruent cues. However, FT was shorter in the 80% reliable condition, relative to the 50% reliable condition, for congruent cues, $p < .001$, but longer for incongruent cues, $p < .001$.

An LME model of FT for trials in which gaze reliability was 100% (Figure 4b) showed no overall difference

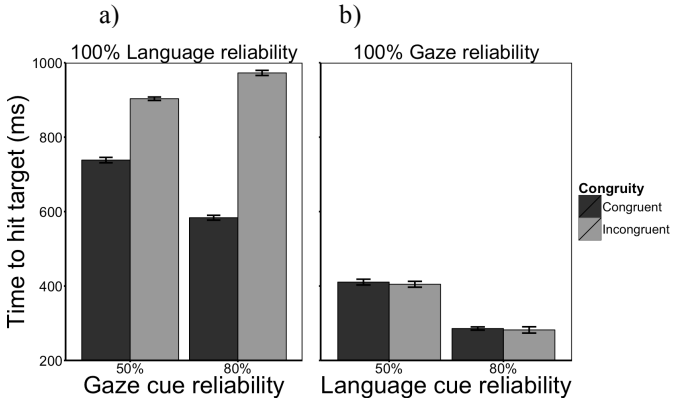

Figure 4: The mean time to fixate target (ms) for participants in congruent and incongruent cue trials for a) sessions with 100% language reliability and 50% and 80% gaze reliabilities and b) sessions with 100% gaze reliability and 50% and 80% language reliabilities. Standard error across trials is shown for each mean.

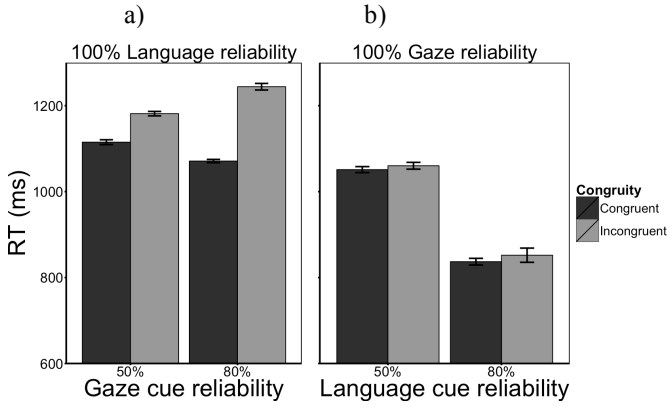

Figure 5: The mean response time (ms) for participants in congruent and incongruent cue trials for a) sessions with 100% language reliability and 50% and 80% gaze reliabilities and b) sessions with 100% gaze reliability and 50% and 80% language reliabilities. Standard error across trials is shown for each mean.

between FT in congruent and incongruent trials, β = 4.870, *SE* = 6.038, *t* = .807, *p* = .424. However, FT was significantly quicker when language reliability was 80% compared to 50%, β = -123.323, *SE* = 55.317, *t* = -2.229, *p* = .045. There was no interaction between these two factors, β = .866, *SE* = 12.232, *t* = .071, *p* = .947.

Finally, we ran LME models of response time (RT). For trials in which language cues were 100% reliable (Figure 5a). Congruent trials were found to have significantly quicker RT scores, β = 115.670, *SE* = 21.120, *t* = -5.478, *p* < .001, and an approaching significant interaction was found between congruity and gaze reliability, β = -100.410, *SE* = 48.480, *t* = -2.071, *p* = .055. There was no overall effect of gaze reliability, β = -10.840, *SE* = 39.680, *t* = -.273, *p* = .761. Post-hoc Tukey tests showed that that incongruent trials had a significantly higher RT when gaze cue reliability was 80% compared to 50%, *p* < 0.001, and that congruent trials had a significantly higher RT when gaze cue reliability was 50% compared to 80% *p* <0.001.

When gaze reliability was 100% (Figure 5b) there was no difference between RT in congruent and incongruent trials, β = -13.445, *SE* = 8.307, *t* = -1.619, *p* = .113. RT was quicker when language reliability was 80% compared to 50%, but this difference was not significant, β = -166.698, *SE* = 101.074, *t* = -1.649, *p* = .108. There was no interaction between these two factors, β = -6.470, *SE* = 13.599, *t* = -.476, *p* = .627.

Discussion

The aim of this study was to investigate how varying the reliability of language cues and gaze cues affects attention and performance in a simple task. We used a target selection task in which the accuracy of both gaze and language cues varied across sessions. In order to gather reliable data each session comprised 480 trials, allowing participants to learn cue reliabilities. Additionally we required that sessions be conducted on different days, to reduce the chance of strategies learned in one session affecting another. Participants generally learned quickly and used the most informative cue provided to inform their response. Gaze reliability had an effect on the time to first fixate the target (FT) and response time (RT), showing that these cues speed up performance when useful. Gaze and language cues had rather different effects on attention and performance when they were the less reliable cue. Language was most disruptive when least reliable, suggesting that it was the favoured cue (Knoeferle & Kreysa, 2012). Gaze cues had a smaller effect when least reliable, supporting the findings of Macdonald and Tatler (2013).

Our analysis of the effects of cue reliability on accuracy shows a clear interaction between the language and gaze cues. Participants strategically followed the instructions of the most reliable cue at their disposal. Neither FT nor RT was significantly influenced by language reliability, but they were affected by gaze reliability, showing that helpful gaze cues benefited performance.

The analysis of congruity effects allowed us to investigate what happens when gaze and language cue different targets, and how this is affected by reliability. When cues were equally, but not entirely reliable (80%), participants followed the language cue more often when the cues were incongruent, suggesting that language is the dominant cue. This finding supports our hypothesis, as well as the results of earlier studies. Macdonald and Tatler (2013) showed that when language and gaze cues were both informative participants used language and ignored gaze, while Knoeferle and Kreysa (2012) found that when processing difficult sentences, participants ignored supportive gaze cues. Both studies suggest a preference for language cues over gaze cues, however it is arguable that in both paradigms the cues are not given equal prominence; in Macdonald and Tatler (2013) gaze cues had to be actively sought out, while verbal instructions were directed toward the participant and in Knoeferle and Kreysa (2012) language comprehension was central to the task and gaze cues were only supportive. In the present study, however, our paradigm gave equal prominence to language and gaze cues, both explicitly (by telling participants that verbal and non-verbal cues were there to help them) and implicitly (by making the reliability of both cues 80%). We have therefore

shown that language cues are preferred over gaze cues when both cues provide the same quality of information.

When the language cues were 100% reliable, gaze stimuli had a larger facilitative effect (when congruent) and detrimental effect (when incongruent) on FT and RT when they were 80% reliable compared to 50% reliable. This suggests that 1) gaze cues are not completely ignored, even when language is 100% accurate and 2) the extent to which they are ignored is inversely related to their perceived reliability. This shows that the gaze cues had more influence when they were more reliable; reliable cues slowed down performance when incongruent with language and sped up performance when congruent.

Overall, the distracting gaze cue results are in line with our hypothesis. When language was 100% accurate, gaze need not (and should not) have been used, however the gaze cues still slowed down performance. These results are typical of a gaze cueing paradigm study (Friesen & Kingstone, 1998). We also found that gaze cues have less influence when they are less reliable, supporting previous findings that the attentional effects of gaze cues are affected by reliability (Hill et al, 2005). Using a real world interaction we have previously found evidence that gaze cues are not sought nor followed when language provides all of the necessary information to complete a task (Macdonald & Tatler, 2013). However, in the present study we have seen that when gaze cues were centrally presented on a screen (i.e. *gaze seeking* is out of the participant's control), gaze following was not inhibited completely by more informative language, but instead gaze following was modulated by the reliability of the cues.

The same effects were not found when language reliability varied alongside 100% reliable gaze cues. FT was significantly slower when language reliability was 50% compared to 80%. We suggest this finding is due to the high rate of incorrect language cues interfering with participants' performance in the task. This effect does not occur with gaze cues because of the relative importance of language compared to gaze in communication; gaze cues are easier to ignore than language. These findings, combined with the preference shown for language cues when presented alongside equally reliable gaze cues, supports our hypothesis that language is the dominant cue.

There was no overall effect of language congruity in these results, suggesting that participants followed the more reliable gaze cues and were uninfluenced by instances of incongruent language. It is surprising to find that the congruity effects of a less reliable gaze cue effect performance, but that the same effects are not found with language cues, particularly given that we have shown evidence that language is the preferred cue. To speculate, it may be because gaze cues were processed faster than language cues and so when the former were entirely reliable, the latter had less time to disrupt processing (RTs were faster for 100% gaze reliability conditions than 100% language reliability conditions).

This experiment investigated the effect of varying gaze and language cue reliability on attention. Participants strategically made use of the most reliable cue to complete the task and favoured language when reliabilities were equal. Our results show that changing the reliability of language affects attention differently than changing the reliability of gaze. Language is the favoured cue and causes overall disruption when unreliable, regardless of congruity, whereas gaze is disruptive when incongruent to more informative language, but only when gaze is more reliable than chance.

Acknowledgments

This paper was supported by an EPSRC Studentship awarded to Ross Macdonald

References

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

Birmingham, E. & Kingstone, A. (2009). Human social attention: A new look at past, present and future investigations. *The Year in Cognitive Neuroscience: Annals of the New York Academy of Sciences, 2009*, 118-140.

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

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.

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

Kuhn, G., & Kingstone, A. (2009). Look away! Eyes and arrows engage oculomotor responses automatically. *Attention, Perception, & Psychophysics, 71* (2), 314-327.

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.

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.

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