Speaker gaze affects utterance comprehension beyond visual attention shifts

Maria Staudte & Matthew W. Crocker
Department of Computational Linguistics, Saarland University
Saarbrücken, Germany

Alexis Heloir & Michael Kipp
DFKI, Embodied Agents Research Group
Saarbrücken, Germany

Abstract

Previous research has shown that listeners follow speaker gaze to mentioned objects in a shared visual environment to ground referring expressions, both for human and robot speakers. What is less clear is whether listeners exploit speaker gaze to infer referential intentions (Staudte & Crocker, 2010), or whether the benefits of gaze can be more simply explained by (reflexive) gaze following (Friesen & Kingstone, 1998). To investigate this issue, we conducted two eye-tracking studies which directly contrast speech-aligned speaker gaze of a virtual agent with a non-gaze visual cue (arrow). Our findings show that speaker gaze is beneficial to listeners only when the order of gaze cues matched the order of mentioned objects in the utterance. Similarly timed arrow cues, however, benefit listeners regardless of the order in which they occur. These findings are consistent with the view that gaze is interpreted as reflecting the speaker's referential intentions, while other visual cues regarding mentioned objects can be exploited more flexibly and strategically.

Keywords: gaze; comprehension; visual attention shift; arrows; referring expressions

Introduction

In face-to-face communication, the speaker's gaze to objects in a shared scene provides the listener with a visual cue to the speaker's focus of (visual) attention (Emery, 2000; Flom, Lee, & Muir, 2007). This potentially offers the listener valuable information to ground and disambiguate referring expressions, to hypothesize about the speaker's communicative intentions and goals and, thus, to facilitate comprehension (Hanna & Brennan, 2007). It is an open question, however, whether this functionality of speaker gaze results simply from its established ability to drive listeners' visual attention, as do other cues as well, or whether gaze uniquely expresses (referential) intentions.

More precisely, there are two levels on which a visual attention shift in response to a speaker's gaze may affect utterance processing (Staudte & Crocker, 2010). On a perceptual level, gaze-following may be considered as (reflexive) visuospatial orienting which increases the visual saliency of the particular target object and/or location in focus (Driver et al., 1999; Friesen & Kingstone, 1998; Langton & Bruce, 1999). On a cognitive level, gaze may *additionally* be understood as a cue to the speaker's referential intentions which elicits expectations about which referent would be mentioned next (Hanna & Brennan, 2007). Previously, these two levels have been identified as the Visual and the Intentional Account, respectively (Staudte & Crocker, 2010). Crucially, the issue whether gaze is processed on both levels – that is, whether the Intentional Account but not the Visual Account alone – offers a satisfying explanation of gaze effects on utterance comprehension, is still under debate. However, recent evidence seems to converge in support of such an assumption (Becchio, Bertone, & Castiello, 2008; Meltzoff, Brooks, Shon, & Rao, 2010; Staudte & Crocker, 2010).

Staudte and Crocker (2010), for instance, synchronized gaze movements of a robot (as one instance of an artificial agent) with its speech in a human-like manner. This was shown to be similarly useful for grounding and resolving spoken references as human gaze (Hanna & Brennan, 2007). Further, Staudte and Crocker (2010) have shown that the order of respective gaze and speech cues is important for efficient comprehension whereas the temporal alignment of those cues is not. That is, when referential gaze cues and the corresponding referring expressions occurred in a coherent linear order, utterance comprehension was facilitated. When this order was reversed, however, gaze did not only *not* help but instead even *slowed* comprehension. In contrast, whether the respective gaze cues occurred one second or five seconds prior to the corresponding referring expression onsets, did not affect the facilitative or disruptive influence of gaze order on comprehension time.

Previous studies have typically manipulated only the validity or credibility of such gaze cues and neglected a direct assessment of the question whether effects on utterance processing are due to shifts in visual attention per se, or whether *speaker gaze* specifically (as opposed to other *exogenous* or even *endogenous* visual cues, Posner, 1980) elicited those attention shifts. To further explore the hypothesis that gaze is indeed interpreted with respect to referential intentions, we adopt an improved experimental design from (Staudte & Crocker, 2010) with a virtual character replacing the robot. We then contrast the influence of gaze and arrows by replacing the gaze cue with an arrow cue, directly comparing the effects of two, possibly different types of visual cues. Specifically, we report supporting evidence from two studies that, firstly, replicate the results on the relevance of gaze cue order for comprehension (Experiment 1) and, secondly, show that other, purely visual cues such as arrows (Experiment 2), induce similar attention shifts as gaze but crucially lack an effect of (inconsistent) order. This supports the hypothesis that gaze does, but arrows do not, elicit inferences about referential intentions.

Exogenous and endogenous cueing

Before describing the experiments, we shall briefly summarize previous findings on different types of (visual) cueing.

It has previously been suggested that gaze-following is a behavior that is applied so reliably that it may be considered automatic. Specifically, studies have shown that people *reflexively* follow stylized gaze cues and other direction-giving cues such as arrows to a target location (e.g. Ristic, Friesen, & Kingstone, 2002). One important issue within this paradigm has been the question whether such gaze cues and arrows, for instance, elicit the same type of attention shift or whether gaze is in some way special (Bayliss & Tipper, 2005; Tipples, 2008). Beyond the reflexive attention shifts mentioned above (also called *exogenous* cueing), people have further been shown to voluntarily orient towards symbolic cues when there is reason to consider these as useful (also called *endogenous* cueing; Posner, 1980). Importantly, both cueing effects have been observed for gaze as well as arrows. That is, when gaze or arrow cues are learned to be counterpredictive (cueing one direction but reliably predicting the target in the opposite direction) they also trigger voluntary attention shifts (see Friesen, Ristic, & Kingstone, 2004; Tipples, 2008; Hanna & Brennan, 2007, for arrows and gaze respectively).

Thus, the reported evidence seems to suggest that reflexive and voluntary orienting applies to both gaze *and* arrows. However, a large body of research has shown that gaze often not only drives visual attention but that it further reveals complex mental states and even intentions (Baron-Cohen, Campbell, Karmiloff-Smith, Grant, & Walker, 1995; Meltzoff et al., 2010). It seems that a whole life time of experiences with gaze has taught people what gaze can reveal about somebody's beliefs, intentions, or emotions, and how useful it may be in various situations (Tomasello & Carpenter, 2007). Thus, the motive to follow gaze may well be qualitatively different from the motive to follow an arrow, for instance, such that arrows may in fact not have identical effects with gaze. The crucial question is what precisely *is* different when performing an attention shift to follow an arrow, compared to following someone's gaze, and whether this difference is measurable.

We hypothesize that while initially a listener may reflexively follow both, gaze and arrows, there are different endogenous motivations for using these cues: In the case of gaze, we hypothesize that the previous experience of its meaningfulness, in particular with respect to intended referents (e.g. Griffin, 2001; Meyer, Sleiderink, & Levelt, 1998), elicits inferences of referential intentions. Thus, a certain order of gaze cues is predicted to elicit expectations for that same order of according speech cues. In contrast, we hypothesize that other visual cues such as arrows, which also direct attention (reflexively and voluntarily), carry no such bias or requirement for a congruent order of cues as they do not lead to inferences of referential intentions. Instead, voluntary orienting towards those cues would occur only if the experimental design and task assigned a temporary benefit to them.

Congruent:	<**s**>	"The star is taller than the <**p**> pyramid."
Reverse:	<**p**>	"The star is taller than the <**s**> pyramid."
Neutral:		"The star is taller than the pyramid."

Figure 1: Sample scene from Experiment 1, with the utterance and congruent/reverse/neutral gaze cues (at pyramid (<**p**>) or at star (<**s**>)).

Experiment 1

In this study, we investigated whether listeners infer referential intentions from agent gaze such that the agent's gaze cues need to be sequentially aligned with corresponding referential speech cues (in the way human gaze is synchronized with produced referring expressions, i.e., preceding the onset of the referring noun by approximately 800ms) in order to be beneficial. Alternatively, agent gaze may be used as a purely visual cue which (reflexively) directs listeners' attention to an object. In the latter case, a "misaligned" sequence of cues may still be beneficial since agent gaze draws attention to mentioned objects in the scene.

We manipulated sequential alignment of gaze and head movement with speech cues to assess the influence of this alignment on comprehension. Specifically, we indirectly measured effects on comprehension by recording response times for utterance validation. The factor "Cue Order" had three levels: The sequence of two referential gaze cues and two referential nouns was either congruent, reverse to each other, or neutral, i.e., straight ahead (Figure 1). Importantly, agent gaze was always directed to mentioned objects only.

Method

Participants Twenty-four native speakers of German, mainly students enrolled at Saarland University, took part in this study (16 females). All participants reported normal or corrected-to-normal vision.

Materials We created 1920x1080 resolution video-clips showing the virtual character *Amber* (Heloir & Kipp, 2009) located behind a table. In each video, there were seven objects on the table, differing in shape and color. *Amber* performed a sequence of head and eye movements consecutively towards two objects in this scene which she also mentioned in a simultaneous utterance, e.g., "The star is taller than the pyramid". The utterance was a synthesized German sentence using the Mary TTS system (Schroeder & Trouvain, 2001).

We manipulated the factor "Cue Order" (congru-

ent/reverse/neutral) so that each item appeared in three conditions. Due to technical reasons, the temporal delay between the onsets of gaze and corresponding speech cues was on average 420 milliseconds for the first noun ("star") and 1030ms for the second noun ("pyramid"). A sample stimulus in all three conditions is depicted by Figure 1. In total, six lists of stimuli were created, accounting for three conditions and their counter-balanced versions. In addition to 24 items, 36 fillers were included such that a total of 60 trials was shown. Fillers frequently contained false utterances to motivate the validation task. The order of item trials was randomized for each participant individually.

Procedure An EyeLink II head-mounted eye-tracker monitored participants' eye movements on a 24-inch monitor. Before the experiment, participants received written instructions about the experiment procedure and task: They were asked to attend to the presented videos and judge whether or not *Amber's* statements were valid with respect to the scene. In order to provide a cover story for this task, participants were further told that the results were used as feedback in a machine learning procedure to improve the agent's performance. The entire experiment lasted approximately 25 minutes.

Analysis Videos were segmented into Interest Areas (IAs). That is, in each video there were labeled regions containing the objects referred to by the first noun (star) and the second noun (pyramid) as well as *Amber's* head. Further, we recorded participant fixations on these regions and report inspection probabilities for the following time windows: GAZE1 stretched from the onset of the initial gaze cue to the onset of the first noun ("star") with a duration of 430ms; N1 contained the first noun and had a mean duration of 386ms; GAZE2 stretched from the onset of the second gaze cue to the onset of the second noun ("pyramid") and was 1,030ms long; N2 contained the second noun and was 385ms long on average. The elapsed time between the second noun onset and the moment of the button press was considered as the response time (RT). Trials were removed when participants had pressed the wrong button (13.4%, Cue Order did not affect accuracy). We further excluded trials as outliers when the response time was $\pm 2.5 * SE$ above or below a participant's mean (1.89 %). Inferential statistics were carried out using mixed-effect models from the lme4 package in R (Baayen, Davidson, & Bates, 2008).

Results

Response Time The mean response times in this experiment are depicted by Figure 2 (error bars show the standard error). For inferential statistics, we log-transformed the response times to obtain normally distributed data. An ANOVA was run on the model fitting the transformed data, as specified in Table 1, and revealed a main effect of Cue Order on response times ($F = 53.42, df = 2$). Table 1 shows the model details and the pairwise comparison between the neutral and the congruent condition, and between the neutral and the reverse condition. In both cases, the t-values as well as p-Values calculated through Monte-Carlo-sampling reveal a significant difference between the levels. Participants were significantly faster in the congruent condition (expressed by the negative coefficient in comparison to the neutral condition) and significantly slower in the reverse condition.

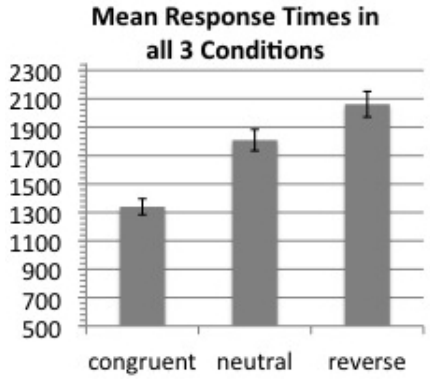

Figure 2: Avg. response times in all three conditions (Exp1).

Table 1: Model fitted to response time data. The last column shows p-Values calculated through Monte-Carlo-sampling.

Predictor	Coeff.	*SE*	t-value	pMCMC
(Intercept, neutral)	7.40	0.067	109.69	<.001
Order-congruent	-0.29	0.039	-7.49	<.001
Order-reverse	0.10	0.042	2.51	<.05

$$Model: log(RT) \sim CueOrder + (1|subject) + (1|item)$$

Eye movements The time curves in Figure 3 plot listeners' fixations towards the star, the pyramid and *Amber's* head as *Amber* looks towards the these objects while also uttering her description. In the top graph (congruent), *Amber's* first gaze movement towards the star is marked by the first, and slightly darker, shaded area prior to the mentioning of the noun "star". The second gaze movement is marked by the second, shaded area prior to mentioning the "pyramid". This pattern is reversed in the reverse condition: The first shaded area marks *Amber's* gaze towards the pyramid before she then mentions the "star". She subsequently looks towards the star (second shaded area) and finally mentions the "pyramid".

The plots clearly show that listeners followed *Amber's* gaze towards the corresponding objects. Already before the onset of the "star" (GAZE1), participant inspections on the star were significantly more likely in the congruent condition than in the reverse ($p < 0.001$) or neutral condition ($p < 0.001$). Similarly, the pyramid was inspected more frequently in the reverse condition compared to congruent ($p < 0.001$) or neutral agent gaze ($p < 0.001$). In the neutral condition, in contrast, inspections on the star and pyramid were equally likely during GAZE1 (probability of 0.03 for both objects) and rose

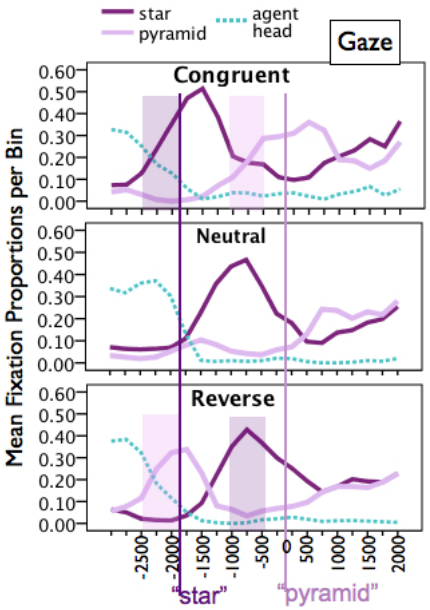

Figure 3: The time curves depict fixations to *Amber's* head and to the star and the pyramid as *Amber's* utterance "The star is taller then the pyramid" unfolds. The plot is aligned to onset of "pyramid", as this marks the most important point of information integration during the multimodal utterance.

more for the star after noun onset (0.13, versus 0.08 for the pyramid). Similar gaze- and speech-following patterns were observed for the time windows GAZE2 and N2.

Discussion

By manipulating Cue Order in the sentence, we created a mismatch between visual and spoken references. This enabled us to observe which reference participants follow initially and how they recover from such a mismatch. The response time data suggest that people found the congruent condition easiest to process and the corresponding eye-movements suggests that this was the case because listeners followed *Amber's* gaze and used it to anticipate the intended next referent. In the reverse condition, participants were slowest which suggests that *Amber's* reversely ordered gaze cues disrupted the comprehension process. The eye-movement data supports the interpretation that listeners infer a referential intention from gaze as there are no signs of recovery from the reversed pattern. That is, even though speaker gaze was obviously relevant (the agent always looked at the two mentioned objects, never at an irrelevant one), we observed looks towards the pyramid mainly during GAZE1 and N1 and hardly before its mentioning, in GAZE2. Even though the reverse condition provides information about both referents of the sentence earlier than the other two conditions (first agent gaze towards pyramid, then mentioning of the "star"), participants were unable to make use of this information and predict the mentioning of the pyramid – but instead were persistently disrupted by the mismatch of *Amber's* referential gaze and speech cues.

Experiment 2

In Experiment 2, we exchanged *Amber's* gaze cue by an arrow appearing above the corresponding object (see Figure 4). This manipulation sought to reveal whether the facilitating and disruptive effects of gaze found in Experiment 1 were caused by the elicited visual attention shifts per se, or whether they were caused because listeners inferred the agent's intention to mention that object from the gaze cue. Given that both gaze and arrow cues (reflexively) direct visual attention in a similar temporal manner, the former hypothesis predicts identical effects on comprehension for arrow and gaze cues whereas the latter hypothesis predicts an adaptation to the utility of the arrow cue (as in the case of counterpredictive cues, for instance). That is, instead of a persistent disruption effect in the reverse condition, it would be expected that a learned association between the arrows/cued objects and the utterance would lead to a beneficial effect of the arrows, regardless of order.

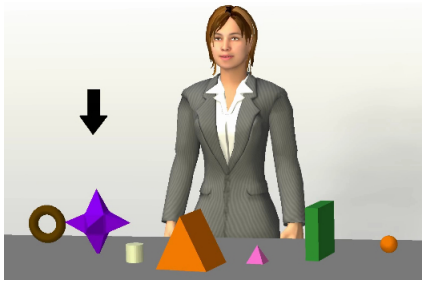

Figure 4: Sample scene from Experiment 2 showing an arrow.

Method

Participants & Procedure Another twenty-four native speakers of German took part in this study (15 females). Again, all reported normal or corrected-to-normal vision. The presence of the arrows was explained to be a cue for *Amber*, to tell her which objects she should talk about, and that sometimes she would not adhere to this. Crucially, we replicated this experiment using an alternative instruction in which participants were told that Amber displayed the arrows to indicate her current interest. This ensured that effects of cue type could not only be attributed to differences in whose intentions the arrows reflected (the experimenter's versus Amber's) and whether they were (im)perfectly valid. Task and Procedure were otherwise identical to Experiment 1.

Materials & Analysis The number and constitution of stimuli was identical to Experiment 1 except for the actual cue. That is, the gaze movement of *Amber* was replaced by an arrow above the respective object for the same onset and duration that *Amber's* gaze would have otherwise identified the object. Consequently, IAs and time windows used in this experiment were identical to Experiment 1 but were extended with the two IAs containing the arrows. Again, trials with

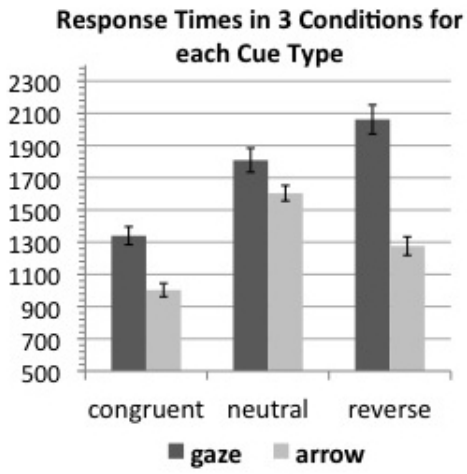

Figure 5: Avg. response times in all three conditions, now in direct comparison between the cue types *gaze* (Exp1) and *arrow* (Exp2).

false responses (8.9%) and outliers (another 2.39%) were removed.

Results

Since the results of this experiment did not qualitatively change with changing the instructions, we report only analyses from the first version in which arrows were introduced as external experimenter cue.

Response Time The mean response times in this experiment are displayed by Figure 5 in direct comparison to the means from Experiment 1. An ANOVA again revealed a main effect of Cue Order on response times ($F = 53.42, df = 2$). The pairwise comparisons are shown in Table 2 and revealed a significant difference between the neutral condition and both the congruent and the reverse condition. This time, however, participants were faster both in the congruent as well as the reverse condition (negative coefficients) compared to the neutral condition.

Table 2: Model fitted to response time data. The last column shows p-Values calculated through Monte-Carlo-sampling.

Predictor	Coeff.	*SE*	t-value	pMCMC
(Intercept, neutral)	1614.24	81.39	19.82	<.001
Order-congruent	-602.11	55.38	-10.87	<.001
Order-reverse	-326.56	57.00	-5.73	<.001

$$Model: log(RT) \sim CueOrder + (1|subject) + (1|item)$$

A combined analysis treating both experiments as a between-subject manipulation of Cue Type (gaze versus arrow), further revealed a main effect of Cue Type ($\chi^2(1) = 9.85, p < .01$) – that is, participants were generally faster in the arrow experiment – as well as an interaction between Cue Type and Cue Order ($\chi^2(2) = 32.49, p < .001$).

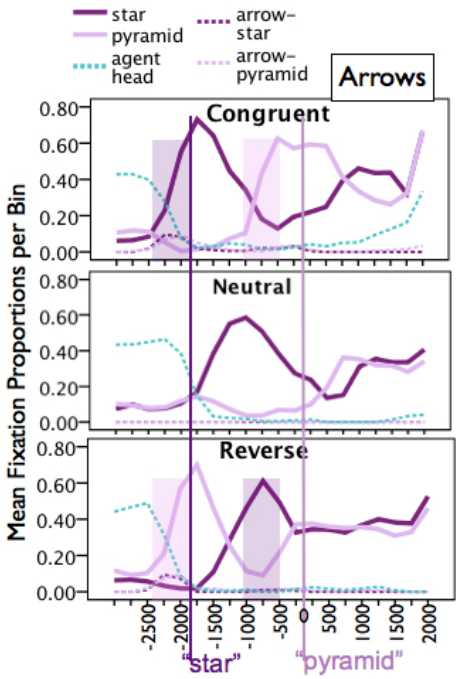

Figure 6: The time curves depict fixations to *Amber's* head, to the star and the pyramid, and to the regions containing the arrows, as *Amber's* utterance "The star is taller then the pyramid" unfolds.

Eye movements The time curves plotted in Figure 6 again show listener fixations on the star, pyramid and *Amber's* head but additionally show fixations on the areas in which the arrows occurred. Crucially, participants hardly looked at the arrow regions and fixation and inspection patterns were indeed surprisingly similar to the ones in Experiment 1. Fixation proportions were generally higher (the scale maximum of the plot is now at 0.8 instead of previously 0.6) showing that participants paid more attention to the objects, in general. Surprisingly, in each trial participants started by fixating *Amber's* head, just like in Experiment 1, even though it never moved. More importantly, there is a difference between the fixation patterns in the reverse conditions of both cue types: While in the case of gaze, listeners rarely looked back at the pyramid before its mention (inspection probability in GAZE2: 0.16), this is not the case for arrows (probability = 0.40). We further conducted a correlation analysis of the actual time of the first fixation back to the pyramid in a time window starting 1,000ms *before* "pyramid" onset. A 0.498 Pearson correlation ($p < 0.001$) was revealed between the time of the first fixation to the pyramid and the response time: The earlier the first fixation happened, the more likely was a short response time. Crucially, the mean first fixation to the pyramid occurred significantly earlier, frequently even before "pyramid" onset, in the arrow study (92ms after noun onset) than in the gaze study (after 662ms, $p < 0.001$).

Discussion

The large inspection probability on the pyramid before its mention (in the reverse arrow condition) suggests that listeners were able to remember and use the earlier arrow cue to the pyramid (even though the agent subsequently mentioned the star) and to predict that this object would be mentioned next. We suggest that for the same reason listeners were faster in the reverse arrow condition than in the neutral condition – in contrast to the reverse gaze condition which disrupted listeners. Significantly less such anticipatory eye-movements to the pyramid in GAZE2 were present in Experiment 1. This suggests that participants detected the task-relevant utility of the arrows and used it even in the reverse condition for predicting the second referent and minimizing response time. Gaze, in contrast, seems to carry a strong bias towards inferring the next intended referent such that participants were unable to use the task-specific utility of the reverse gaze cues. Thus, they did not predict the second referent in the reverse condition, resulting in longer response times. Further evidence for this adaptiveness to cue utility in the case of arrows, but not gaze, is provided by a block analysis: In Experiment 1, there was a main effect of Block ($\chi^2(1) = 4.26, p < .05$), showing that participants became faster in general, but crucially there was no interaction of Block and Cue Order. In Experiment 2, however, there was a main effect ($\chi^2(1) = 11.00, p < .001$) as well as an interaction ($\chi^2(2) = 11.64, p < .01$) carried mainly by the speed up in the reverse condition: From a mean of 1,477.35ms to 1,123.28ms. This suggests that participants improved in exploiting the predictive power of arrows, but not gaze, over time.

Conclusion

The presented findings support the position that listeners use speaker gaze to infer referential intentions and predict the spoken references to occur in similar linear order. Further, the evidence from Experiment 1 suggests that this inference is drawn almost automatically so that listeners cannot easily adapt to the counterpredictive utility of the agent's (reverse) gaze cues. While Hanna and Brennan (2007) observed that their participants did adapt to and use the spatial counterprectiveness of speaker gaze, this was found in a blocked design, giving participants sufficient training to adapt to this situation. Our results, in contrast, suggest that listeners originally shift their attention in response to both gaze and arrow cues in a similar manner while only adapting to the task-specific (and temporally counterpredictive) utility of arrows, applying this spontaneously created association to predict referents *independent* of their order of mention. These results provide evidence for a bias of using gaze to infer referential intentions and a lack thereof in using arrows. It remains an open question, however, whether the endogenous motivation for using gaze in this way is based on previously formed probabilistic models of co-occurrence (referential gaze often preceding the mentioning of an object) or on a qualitative model of the function of gaze and the speaker's intentional states.

Acknowledgments

The research reported of in this paper was supported by the "Multimodal Computing and Interaction" Cluster of Excellence at Saarland University.

References

Baayen, R., Davidson, D., & Bates, D. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390-412.

Baron-Cohen, S., Campbell, R., Karmiloff-Smith, A., Grant, J., & Walker, J. (1995). Are children with autism blind to the mentalistic significance of the eyes? *British Journal of Developmental Psychology*, *13*, 379-398.

Bayliss, A., & Tipper, S. (2005). Gaze and arrow cueing of attention reveals individual differences along the autism spectrum as a function of target context. *British Journal of Psychology*, *96*, 95-114.

Becchio, C., Bertone, C., & Castiello, U. (2008). How the gaze of others influences object processing. *Trends in Cognitive Science*, *12*, 254-258.

Driver, J., Davis, G., Ricciardelli, P., Kidd, P., Maxwell, E., & Baron-Cohen, S. (1999). Gaze Perception Triggers Reflexive Visuospatial Orienting. *Visual Cognition*, *6*, 509-540.

Emery, N. (2000). The eyes have it: the neuroethology, function and evolution of social gaze. *Neuroscience and Biobehavioral Reviews*, *24*, 581-604.

Flom, R., Lee, K., & Muir, D. (Eds.). (2007). *Gaze-Following: Its Development and Significance*. Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

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

Friesen, C., Ristic, J., & Kingstone, A. (2004). Attentional Effects of Counterpredictive Gaze and Arrow Cues. *Journal of Experimental Psychology: Human Perception and Performance*, *30*(2), 319-329.

Griffin, Z. M. (2001). Gaze durations during speech reflect word selection and phonological encoding. *Cognition*, *82*, B1-B14.

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

Heloir, A., & Kipp, M. (2009). EMBR - A Realtime Animation Engine for Interactive Embodied Agents. In *Proceedings of the 9th International Conference on Intelligent Virtual Agents (IVA '09)*. Sprinter.

Langton, S. R., & Bruce, V. (1999). Reflexive Visual Orienting in Response to the Social Attention of Others. *Visual Cognition*, *6*, 541-567.

Meltzoff, A. N., Brooks, R., Shon, A. P., & Rao, R. P. N. (2010). Social robots are psychological agents for infants: A test of gaze following. *Neural Networks*, *23*, 966-972.

Meyer, A., Sleiderink, A., & Levelt, W. (1998). Viewing and naming objects: Eye movements during noun phrase production. *Cognition*, *66*, B25-B33.

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

Ristic, J., Friesen, C. K., & Kingstone, A. (2002). Are eyes special? It depends on how you look at it. *Psychonomic Bulletin & Review*, *9*, 507-513.

Schroeder, M., & Trouvain, J. (2001). The German Text-to-Speech Synthesis System MARY: A Tool for Research, Development and Teaching. In *4th isca workshop on speech synthesis.* Blair Atholl, Scotland.

Staudte, M., & Crocker, M. W. (2010). When Robot Gaze Helps Human Listeners: Attentional versus Intentional Account. In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd Annual Conf. of the Cognitive Science Society.* Portland, OR.

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

Tomasello, M., & Carpenter, M. (2007). Shared Intentionality. *Developmental Science*, *10*, 121-125.