

Why does the gaze of others direct visual attention?

Paul E. Downing, Chris M. Dodds, and David Bray

School of Psychology, University of Wales, Bangor, UK

Viewing another person directing his or her gaze can produce automatic shifts of covert visual attention in the same direction. This holds true even when the task-relevant target is much more likely to occur at the uncued location. These findings, along with other evidence, have been taken to suggest that gaze represents a “special” stimulus—the foundation of a social cognition system that can make inferences about the mental states of other people. However, gaze-driven cueing effects could simply be due to spatial compatibility between cue and target. We compared the attentional effects of gaze shifts to a face with the tongue extended laterally to the left or right. When tongue direction was a nonpredictive cue, we found cueing effects from tongues that were indistinguishable from those produced by gaze. However, in contrast to previous findings with gaze, tongue cues did not overcome a validity manipulation in which targets were four times more likely to appear at the uncued location. We conclude that simple attentional cueing effects from gaze may be better explained by spatial compatibility, and that more complex, unique features of cueing from gaze may be better indices into perceptual systems specialized for social cognition.

One aim of social cognitive neuroscience has been to understand how we infer the mental states of other people. Particular attention has been given to faces, which can indicate emotional states and intentions through facial expressions (Ekman, Friesen, & Ellsworth, 1972), blushing (Leary, Britt, Cutlip, & Templeton, 1992), and direction of gaze (Emery, 2000; Langton, Watt, & Bruce, 2000). Several researchers have recently focused on gaze as playing a key role in the development of the “theory of mind”. For example, Baron-Cohen (1995) has proposed an “eye direction detector” module, a primitive cognitive system that detects the direction of another person’s gaze. This system is proposed to support development of the ability to infer what that person sees, and, by extension, the ability to represent other people as independent mental agents.

Please address all correspondence to: P. Downing, School of Psychology, Centre for Cognitive Neuroscience, University of Wales, Bangor, LL57 2AS, UK. E-mail: p.downing@bangor.ac.uk

We thank S. Hayward, K. Silk, G. Williams, and J. Miller for helpful comments and assistance with data collection. This research was supported by BBSRC grant 5/S15293 to P. Downing.

Several converging techniques have been used to study gaze perception mechanisms. Functional neuroimaging studies (Haxby, Hoffman, & Gobbini, 2002; Puce, Allison, Bentin, Gore, & McCarthy, 1998) show that regions of the superior temporal sulcus (STS) become activated by faces changing their direction of gaze. Similar activations are seen to changes of facial expression, suggesting a general role for this region in processing dynamic aspects of the face. Lesion studies also implicate this area bilaterally in perception of gaze direction (Campbell, Heywood, Cowey, Regard, & Landis, 1990; but see Eacott, Heywood, Gross, & Cowey, 1993). And psychophysical studies (Ricciardelli, Baylis, & Driver, 2000) demonstrate high sensitivity to gaze direction that depends on the relative contrast of the sclera and pupil, suggesting a specialized mechanism tuned to particular visual features.

A number of recent studies have shown that viewing the diverted gaze of another person triggers a reflexive attentional shift in the same direction (Driver, Davis, Ricciardelli, Kidd, Maxwell, & Baron-Cohen, 1999; Langton & Bruce, 1999; Ristic, Friesen, & Kingstone, 2002). In the typical procedure, a cue face is presented centrally, with gaze directed left or right. A target item requiring a speeded response then appears either to the left or right of the display. Target items are discriminated or detected more rapidly when they appear in a location congruent with the direction of gaze of the central face. These cueing effects are found even when the direction of gaze is not predictive of target location. This effect has been demonstrated with both schematic faces and face photographs. A recent ERP study (Schuller & Rossion, 2001) has shown that the N1 and P1 components triggered by target stimuli in a similar procedure are enhanced when they appear in a congruent location, relative to an incongruent location. Comparative studies (Emery, Lorincz, Perrett, Oram, & Baker, 1997) suggest that other primates such as chimpanzees and macaques spontaneously follow the gaze of conspecifics as well. And finally, infants show an early sensitivity to gaze (Farroni, Csibra, Simion, & Johnson, 2002; Vecera & Johnson, 1995), and a similar shift in attention following the direction of gaze of adults (Hood, Willen, & Driver, 1998).

EXPERIMENT 1

A key feature of the procedure outlined above is that the direction of gaze of the face is the only salient change from trial to trial. Perhaps subjects automatically encode the cue faces in terms of “left” and “right” simply because this is the only dimension that varies in the stream of stimuli (see also Tipples, 2002). According to this account, changes in any facial feature that induced left/right coding in a similar way should produce similar attentional cueing costs and benefits.

To this end, we compared attentional cueing effects from gaze, replicating previous research, to cueing from left or right projections of the tongue. Clearly

there can be no specialized visual module for detecting the direction of others' tongues. If congruency between tongue direction and target location speeds target discrimination, it would suggest that previously seen attentional effects of gaze can be more parsimoniously explained by left–right coding of the cue faces. In contrast, if attentional cueing is seen only with gaze, this would provide strong evidence that there is something “special” about gaze-driven cueing effects, above and beyond spatial compatibility effects.

In two otherwise identical versions of the experiment, we manipulated the duration of the tongue or gaze cue: in Experiment 1A this was fixed at 507 ms, and in Experiment 1B it could be either 107 or 707 ms.

Methods

Participants. Thirty-two members of the Bangor community participated in Experiment 1A, and forty in Experiment 1B. Participants were assigned to “gaze” or “tongue” groups in an alternating fashion, and were compensated with either £5 or with course credit.

Stimuli. The experimental stimuli are illustrated in Figure 1. A single photograph of a male face was used as the basis for all stimuli, so that any image asymmetries unrelated to the experimental manipulation would be identical across conditions. This base image was used as a “neutral” alerting signal in all conditions. Cue stimuli were generated by replacing either the eye or mouth regions of the base image. Thus only these regions changed when the neutral face was replaced with a cue face.

The face image was approximately 8° visual angle in width and 12° in height. The faces were presented with the tip of the nose at the centre of the screen. The target letters “T” and “L” were used for the discrimination task. These were presented at the horizontal midline of the screen, approximately 6° from the centre, so that they would be equidistant from the face parts used as cues in the two experimental groups. The target letters were approximately $1.0^\circ \times 1.2^\circ$ in size.

Design. In both experiments, cue type (tongue or eyes) was manipulated between subjects, and congruency between cue and target location (congruent or incongruent) was manipulated within subjects. In Experiment 1A, the direction of the cue (left or right), the location of the target (left or right), and the target type (T or L) were fully counterbalanced in each block of 16 trials. The same design was used for Experiment 1B, with the addition of a fourth orthogonal variable, cue duration (107 or 707 ms). Seven blocks were tested in Experiment 1A and eleven in Experiment 1B. In the analysis of both experiments, data from the first block was discarded.

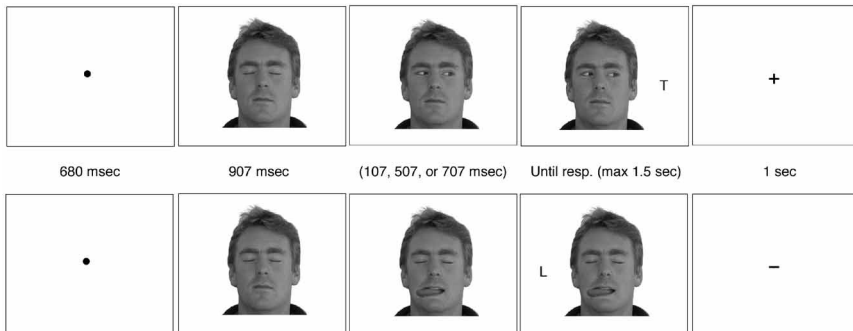


Figure 1. Illustration of the procedure for Experiments 1A and 1B. Timing for each frame is shown in the centre row. Stimuli for the tongue and eye conditions were exactly matched except on the critical stimulus dimensions.

Procedure. The experimental procedure is illustrated in Figure 1. Each trial began with a central fixation point. This was replaced after 680 ms by a “neutral” face with eyes and mouth closed. After 907 ms, the cue was presented: Either the eyes opened and looked left or right, or the mouth opened and the tongue extended to the left or right. In Experiment 1A the duration of the cue was fixed at 507 ms. In Experiment 1B, the cue duration was either 107 ms or 707 ms. After this delay elapsed, the target letter was presented to the left or right of the face. Participants were required to report the identity of the target, pressing the “up” arrow on the keyboard for a T, and the “down” arrow for an L. The response keys were aligned with the centre of the display to reduce compatibility effects on response times. Participants were encouraged to respond quickly and accurately. They were also informed that the face appearing on each trial was not relevant to the task, and that they were simply to report the identity of the target letter.

Responses slower than 1.5 s were treated as errors. Each trial ended with an accuracy feedback signal (a “+” or a “–”), which was presented for 1 s. Participants then pressed the space bar to initiate the next trial, which began after a 1 s delay.

Results

Overall accuracy was very high ($M = 97.0\%$ in Experiment 1A, $M = 97.3\%$ in Experiment 1B). Analysis of response time (RT) data included only trials in which the response was correct. Median RTs were computed for each subject in each condition (see Figure 2) in order to reduce the effects of outliers.

In Experiment 1A, only the main effect of congruency was significant, $F(1, 30) = 19.6$, $MSE = 582$, $p < .001$ (see Figure 2, top panel). RTs were faster when the cue was congruent with the target location ($M = 530$ ms) than when the

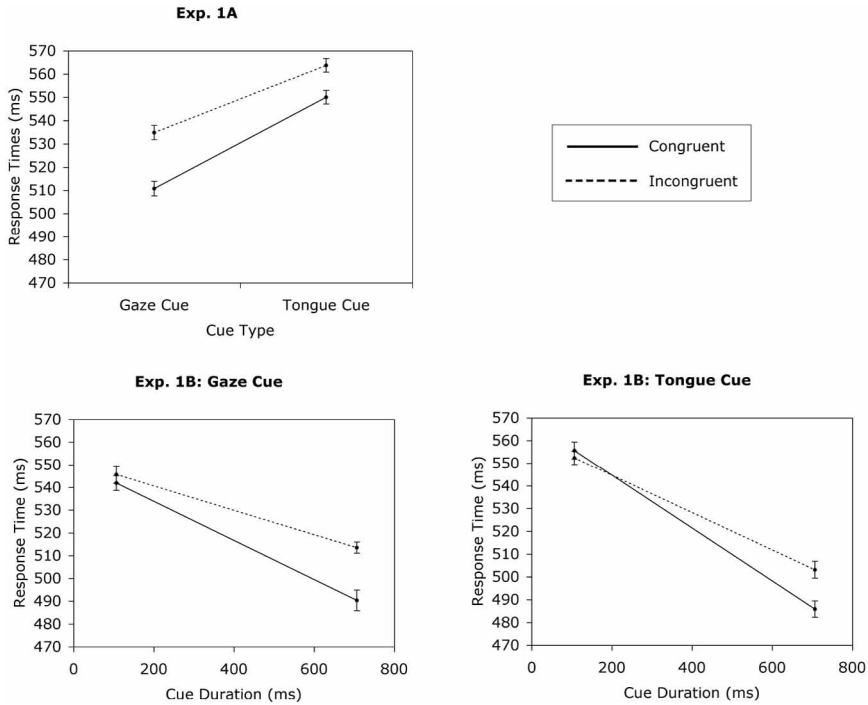


Figure 2. Results of Experiments 1A (top panel) and 1B (bottom panels). In Experiment 1A cue duration was held constant, so means of median RTs are plotted against experimental group and cue–target relationship (see legend). For Experiment 1B, means of median RTs are plotted for each group, as a function of cue duration and cue–target relationship. Error bars represent the standard error after adjusting for between-subjects variance (Loftus & Masson, 1994).

cue was incongruent ($M = 549$ ms). This congruency effect did not depend on whether the eyes or the tongue were used as the cue: The interaction of cue type with congruency was not significant, $p > .20$. Simple effects tests showed that the congruency effect was significant within both the “gaze” group, $F(1, 15) = 15.0$, $MSE = 619$, $p < .01$, and the “tongue” group, $F(1, 15) = 5.5$, $MSE = 546$, $p < .05$.

In Experiment 1B (see Figure 2, lower panels), there were significant main effects of congruency, $F(1, 38) = 11.0$, $MSE = 380$, $p < .01$, and cue duration, $F(1, 38) = 265.9$, $MSE = 386$, $p < .001$. The interaction of group (eye or tongue) and cue duration was significant, $F(1, 38) = 8.1$, $MSE = 386$, $p < .01$, as was the interaction of cue duration and congruency, $F(1, 38) = 16.4$, $MSE = 244$, $p < .001$. Critically, the three-way interaction of group, cue duration, and congruency did not approach significance, $F < 1$. Follow-up analyses showed a Congruency \times Cue duration interaction for both the “gaze” group, $F(1, 19) =$

9.0, $MSE = 210$, $p < .01$, and the “tongue” group, $F(1, 19) = 7.6$, $MSE = 278$, $p < .05$. In both groups there was no effect of the cue at 107 ms, both $F_s < 1$. In contrast, both groups showed a significant congruency effect at 707 ms; eyes: $F(1, 19) = 15.4$, $MSE = 347$, $p < .001$; tongue: $F(1, 19) = 9.2$, $MSE = 325$, $p < .01$.

Finally, we conducted a combined analysis of both studies, with experiment (1A or 1B), congruency (congruent or incongruent), and cue type (eyes or tongue) as factors. Only the main effect of congruency was significant, $F(1, 68) = 30.4$, $MSE = 235$, $p < .001$. The congruency effect was not modulated by whether gaze or tongue extension were used as the cue, $p > .1$.

EXPERIMENT 2

A previous report by Driver et al. (1999) showed that cueing effects from gaze were sufficiently powerful to overcome the endogenous effects of cue predictability. When targets were four times more likely to appear in an uncued location than a cued location, RTs were nonetheless faster to targets at the cued location. This suggests that gaze direction is a powerful, bottom-up cue, relatively insensitive to top-down modulation. In this study, we replicated the procedure of Driver et al. in one group of subjects. In a second group, the same paradigm was used, except that tongue direction rather than gaze served as a cue. If tongue cues are also insensitive to probability manipulations, it would further suggest that cueing effects from gaze are due to spatial compatibility, and that gaze may not be a “special” visual stimulus.

Methods

Participants. Thirty-four members of the Bangor community participated in Experiment 2. They were compensated with either £5 or with course credit.

Stimuli. Apparatus and stimuli were as in Experiment 1.

Design and procedure. Five blocks of 60 trials were tested; the first was discarded as practice. Within each block 12 trials were congruent (6 left and 6 right cues) and 48 trials were incongruent (24 left and 24 right cues). The target type (T or L) was selected randomly on each trial. Timing parameters were the same as in Experiment 1, except that, following Driver et al. (1999), the duration of the cue was 307 ms. Also, following Driver et al., participants were reminded at the beginning of each block that the target letter was “four times as likely to appear on the side away from where the tongue points”. In other respects the procedure matched that of Experiment 1. Seventeen subjects were tested with gaze cues, and seventeen with tongue cues.

Results and discussion

Overall accuracy was 98% correct in the tongue condition and 97% in the gaze condition. Analysis of variance on median correct response times (see Figure 3) showed no significant main effect of cue type (gaze or tongue), $F < 1$, and no significant main effect of congruency between cue and target locations, $F < 1$. However, the interaction of these two variables approached significance, $F(1, 32) = 3.7$, $p = .06$. This interaction was significant in the equivalent ANOVA on mean response times, $F(1, 32) = 4.3$, $p < .05$. In the tongue cue condition, responses were faster at the incongruent (but more likely) location, while in the gaze condition, responses were faster at the congruent (but less likely) location, consistent with Driver et al. (1999).

These results show that the cueing effects generated by tongue projection seen in Experiment 1 are not sufficiently automatic to overcome the top-down effects of target probability. This is in contrast to cueing from gaze, which (replicating Driver et al., 1999) is sufficiently automatic and powerful to overcome the 4:1 probability manipulation.

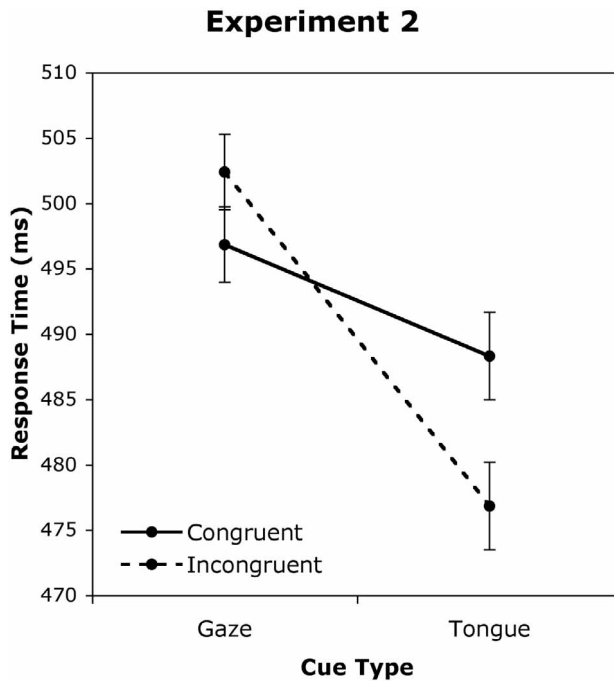


Figure 3. Results of Experiment 2. When gaze direction serves as the attentional cue, target processing is facilitated at the congruent location, even though the target is much more likely at the incongruent location (Driver et al., 1999). In contrast, when attention is cued with tongue position, target processing is facilitated at the incongruent (but more likely) location.

GENERAL DISCUSSION

Spatial congruency between a gaze shift in an image of the human face and the location of a target speeds discrimination of that target. Congruency benefits of this type are typically attributed to an automatic shift of spatial attention in the direction of the cue. Previously, cueing effects from gaze have been taken as evidence for a specialized neural mechanism that computes the direction of attention of other people. As noted above, some researchers have gone further to propose that this system forms the root of a more general cognitive system that interprets the actions of others in order to infer their mental states, beliefs, and intentions.

However, in the present study we found no difference between shifts of attention elicited by nonpredictive gaze shifts and extensions of the tongue. Although numerically somewhat smaller, cueing effects from the tongue did not differ significantly from gaze-elicited effects, even in a combined analysis with 72 subjects. The time course of cueing effects for the two types of cue also did not differ: As in previous work (Driver et al., 1999), we found that 100 ms is insufficient time for congruency effects to develop, and this held for tongue cues as well as gaze cues. We also note that while the two kinds of cueing stimuli used here were controlled at the conceptual level, in that both indicate either the left or right of the display, they were not controlled at the level of visual features. There may be subtle differences between the eye and tongue cues in contrast or salience that contribute to the cueing effects they produce.

In contrast, the tongue condition did not produce automatic shifts of attention sufficient to overcome a 4:1 probability in favour of the target appearing at the uncued location. Accordingly, we suggest that it is the capacity to overcome top-down biases that makes gaze a “special” stimulus, above and beyond the spatial compatibility effects also seen for simpler stimuli (Hommel, Pratt, Colzato, & Godijn, 2001; Tipples, 2002). In a similar vein, Ristic et al. (2002) found apparently reflexive orienting to gaze as well as to simple arrow cues. However, in a split-brain patient (Kingstone, Friesen, & Gazzaniga, 2000) arrow cues produced cueing bilaterally, while gaze cues did so only in the hemisphere specialized for face processing. These authors propose that cueing effects from gaze may be subserved by different brain mechanisms than those that support other cues.

A growing body of evidence has now been gathered in support of cognitive systems specialized for human gaze perception (Baron-Cohen, 1995; Haxby et al., 2002). One main thread of evidence has come from orienting studies like those described here. Recent work, including the present study, now suggests that simple orienting effects may not highlight the unique properties of gaze processing. In general, we suggest that secondary characteristics of gaze cueing, such as resistance to top-down biases, will prove to be a more useful cognitive signature of specialized gaze detection systems.

REFERENCES

- Baron-Cohen, S. (1995). *Mindblindness: An essay on autism and theory of mind*. Cambridge, MA: MIT Press.
- Campbell, R., Heywood, C. A., Cowey, A., Regard, M., & Landis, T. (1990). Sensitivity to eye gaze in prosopagnosic patients and monkeys with superior temporal sulcus ablation. *Neuropsychologia*, *28*, 1123–1142.
- Driver, J., Davis, G., Ricciardelli, P., Kidd, P., Maxwell, E., & Baron-Cohen, S. (1999). Gaze perception triggers visuospatial orienting by adults in a reflexive manner. *Visual Cognition*, *6*, 509–540.
- Eacott, M. J., Heywood, C. A., Gross, C. G., & Cowey, A. (1993). Visual discrimination impairments following lesions of the superior temporal sulcus are not specific for facial stimuli. *Neuropsychologia*, *31*, 609–619.
- Ekman, P., Friesen, W. V., & Ellsworth, P. (1972). *Emotion in the human face*. Elmsford, NY: Pergamon Publishing Co.
- Emery, N. J. (2000). The eyes have it: The neuroethology, function and evolution of social gaze. *Neuroscience and Biobehavioral Reviews*, *24*, 581–604.
- Emery, N. J., Lorincz, E. N., Perrett, D. I., Oram, M. W., & Baker, C. I. (1997). Gaze following and joint attention in rhesus monkeys (*Macaca mulatta*). *Journal of Comparative Psychology*, *111*, 286–293.
- Farroni, T., Csibra, G., Simion, F., & Johnson, M. H. (2002). Eye contact detection in humans from birth. *Proceedings of the National Academy of Sciences USA*, *99*, 9602–9605.
- Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2002). Human neural systems for face recognition and social communication. *Biological Psychiatry*, *51*, 59–67.
- Hommel, B., Pratt, J., Colzato, L., & Godijn, R. (2001). Symbolic control of visual attention. *Psychological Science*, *12*, 360–365.
- Hood, B. M., Willen, J. D., & Driver, J. (1998). Adults' eyes trigger shifts of visual attention in human infants. *Psychological Science*, *9*, 131–134.
- Kingstone, A., Friesen, C. K., & Gazzaniga, M. S. (2000). Reflexive joint attention depends on lateralized cortical connections. *Psychological Science*, *11*, 159–165.
- Langton, S. R., Watt, R. J., & Bruce, V. (2000). Do the eyes have it? Cues to the direction of social attention. *Trends in Cognitive Sciences*, *4*, 50–59.
- Langton, S. R. H., & Bruce, V. (1999). Reflexive social orienting. *Visual Cognition*, *6*, 541–567.
- Leary, M. R., Britt, T. W., Cutlip, W. D. II, & Templeton, J. L. (1992). Social blushing. *Psychological Bulletin*, *112*, 446–60.
- Loftus, G. R., & Masson, M. E. J. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin and Review*, *1*(4), 476–490.
- Puce, A., Allison, T., Bentin, S., Gore, J. C., & McCarthy, G. (1998). Temporal cortex activation in humans viewing eye and mouth movements. *Journal of Neuroscience*, *18*, 2188–2199.
- Ricciardelli, P., Baylis, G., & Driver, J. (2000). The positive and negative of human expertise in gaze perception. *Cognition*, *77*, B1–14.
- Ristic, J., Friesen, C. K., & Kingstone, A. (2002). Are eyes special? It depends on how you look at it. *Psychonomic Bulletin and Review*, *9*, 507–513.
- Schuller, A. M., & Rossion, B. (2001). Spatial attention triggered by eye gaze increases and speeds up early visual activity. *Neuroreport*, *12*, 2381–2386.
- Tipples, J. (2002). Eye gaze is not unique: Automatic orienting in response to uninformative arrows. *Psychonomic Bulletin and Review*, *9*, 314–318.
- Vecera, S. P., & Johnson, M. H. (1995). Gaze detection and the cortical processing of faces: Evidence from infants and adults. *Visual Cognition*, *2*, 59–87.

Manuscript received December 2002
Revised manuscript received June 2003

