Goldsmiths Research Online

Goldsmiths Research Online (GRO)
is the institutional research repository for
Goldsmiths, University of London

Citation

Kuhn, Gustav; Vacaityte, Ieva; D'Souza, Antonia D. C; Millett, Abbie and Cole, Geoff G.. 2018. Mental states modulate gaze following, but not automatically. Cognition, 180, pp. 1-9. ISSN 0010-0277 [Article]

Persistent URL

https://research.gold.ac.uk/id/eprint/23466/

Versions

The version presented here may differ from the published, performed or presented work. Please go to the persistent GRO record above for more information.

If you believe that any material held in the repository infringes copyright law, please contact the Repository Team at Goldsmiths, University of London via the following email address: gro@gold.ac.uk.

The item will be removed from the repository while any claim is being investigated. For more information, please contact the GRO team: gro@gold.ac.uk

Goldsmiths
UNIVERSITY OF LONDON

Mental states modulate gaze following, but not automatically.

Gustav Kuhn[1], Ieva Vacaityte[2], Antonia D. C. D'Souza[3], Abbie Millett, & Geoff G. Cole[3]

1) Department of Psychology, Goldsmiths, University of London, UK

2) Department of Psychology, Brunel University

3) Centre for Brain Science, University of Essex, UK

Correspondence should be addressed to:

Geoff G. Cole
Centre for Brain Science
University of Essex
Wivenhoe Park
Colchester
CO4 3 SQ
E-mail: ggcole@essex.ac.uk
Tel: 01206 872331
Fax: 01206 873801

Abstract

A number of authors have suggested that the computation of another person's visual perspective occurs automatically. In the current work we examined whether perspective-taking is indeed automatic or more likely to be due to mechanisms associated with conscious control. Participants viewed everyday scenes in which a single human model looked towards a target object. Importantly, the model's view of the object was either visible or occluded by a physical barrier (e.g., briefcase). Results showed that when observers were given five seconds to freely view the scenes, eye movements were faster to fixate the object when the model could see it compared to when it was occluded. By contrast, when observers were required to rapidly discriminate a target superimposed upon the same object no such visibility effect occurred. We also employed the barrier procedure together with the most recent method (i.e., the ambiguous number paradigm) to have been employed in assessing the perspective-taking theory. Results showed that the model's gaze facilitated responses even when this agent could not see the critical stimuli. We argue that although humans do take into account the perspective of other people this does not occur automatically.

Introduction

We often gaze towards locations that are looked at by others, and this form of social attention is an essential part of human interaction and cognition in general. At the centre of this orienting mechanism is the need to know what others are looking at, a process that involves the computation of another person's mental state, i.e., Theory of Mind (ToM). Although early work did consider ToM mechanisms in this so-called gaze following, later social attention workers tended to conceive gaze-induced attentional behaviour as a bottom-up process, rather than involving higher mechanisms (Driver, et al., 1999; Friesen & Kingstone, 1998; Ricciardelli, Bricolo, Aglioti, & Chelazzi, 2002). More recently, a number of authors have explicitly suggested that gaze following is influenced by what the gazer can see and that ToM forms an essential component of gaze cueing (e.g., Teufel, Fletcher, & Davis, 2010). Some authors have even argued that the computation of what others see occurs spontaneously (Samson, Apperly, Braithwaite, Andrews, & Scott, 2010).

Mechanisms associated with gaze following are typically investigated using some variant of a paradigm in which participants are asked to respond to targets that either appear in locations looked at by another agent (i.e., 'valid' trials), or they appear elsewhere (i.e., 'invalid' trials; Driver, et al., 1999; Friesen & Kingstone, 1998). Response times (RTs) are generally shorter on valid compared with invalid trials, an effect that has been observed for both manual responses (Frischen, Bayliss, & Tipper, 2007) as well as saccadic eye movements (Kuhn & Benson, 2007; Ricciardelli et al., 2002). To examine whether ToM processes modulate gaze following, Teufel, Alexis, Clayton, and Davis (2010) used a modified version of this task in which the gazing agent wore mirrored goggles. Participants were informed that the goggles were either transparent or opaque, thereby manipulating whether the agent could see the targets or

not. Results revealed a larger gaze cueing effect when participants were informed that the agent could see, thus supporting the view that gaze following is modulated by mental state attribution (see also Teufel et al., 2009).

Although the above results have been taken as evidence that gaze cueing *can* be modulated by mental states, some authors have argued that humans spontaneously compute the perspective of others. This view has come from results obtained in the 'dot perspective' paradigm in which participants are presented with an image of a room that contains an avatar who looks either towards the left or the right hand wall (e.g., Samson et al., 2010; Santiesteban, Catmur, Hopkins, Bird, & Heyes, 2014). A number of dots are pinned to either the left, the right, or both walls, and participants are asked to make judgments about the number of dots that are either visible to them or visible to the avatar. The most interesting finding from this procedure is that when making own-perspective judgements, participants make slower responses if the number of dots seen by the avatar does not match that seen by the participant; so-called altercentric intrusion. Samson et al. argued that this occurs because the avatar's perspective is computed, leading to the interference, and that this process is 'spontaneous'. In a later article, Surtees and Apperly (2012) stated that the process is 'automatic'. Samson et al. (2010) also suggested that the processes involved in this effect are similar to those involved in generating gaze following in the gaze cueing effect (described above).

A central challenge to the mental state theory of these gaze-induced effects has come from a series of experiments by Cole and colleagues (Cole, Atkinson, Le, & Smith, 2016; Cole, Smith, & Atkinson, 2015; Cole, Atkinson, D'Souza, & Smith, 2017; see also Langton, 2018) who adopted a procedure often employed in animal and infant ToM research (e.g., Moll & Tomasello, 2004). A physical barrier placed in the line of sight between the gazing agent and the target renders the target non-visible to the gazer.

For instance, Hare, Call, and Tomasello, (2001) showed that a subordinate chimpanzee knows whether a dominant chimpanzee can see a food item based on whether the latter's view of the food is obscured by a barrier or not. Since a gazer cannot see a target under a barrier condition, Cole et al. reasoned that any gaze cueing-like effect observed when a target is not visible to the gazer cannot be due to the gazer's visual perspective driving the gaze cueing effect. In a series of experiments, including one in which a physically present person acted as the cue, the gazing agent induced strong cueing effects. Importantly, Cole et al. found that these effects were not influenced by whether the target was visible or not, thus challenging the notion that gaze effects are modulated by mental states. In a follow-up study, Cole et al. (2016) found the same pattern of results when the barrier method was employed in the dot perspective task. That is, automatic perspective-taking-like data were observed when the avatar could not see any dots due to the location of a barrier.

The findings of Cole et al. are however in contrast to other recent work that has also employed the barrier method. Using a variant of the dot perspective task, Baker, Levin, and Saylor (2016) reported that dot judgements *were* found to be influenced by whether the avatar could see the targets or not (see also Morgan, Freeth, & Smith, 2018). Furthermore, the challenge to the theory that ToM influences gaze following does not concur with the common observation that we, as social beings, often find ourselves following another person's gaze precisely because of an explicit mental state attribution. Many of us have often said to ourselves "I wonder what she is looking at" before trying to determine what the viewer is observing. This is a clear example of visual perspective modulating gaze following. The real issue may therefore be whether the process in which ToM modulates gaze following can occur automatically.

The primary aim of the present work was to examine the question of whether humans do indeed compute the perspective of other individuals. In four experiments we adopted the visibility manipulation described above in which a gazing agent either sees the target stimuli or does not. Furthermore, our experiments were particularly concerned with the theory that not only does such perspective computation occur but that it does so automatically. As we briefly review in the General Discussion, the notion of automaticity has been somewhat problematic, with different authors suggesting a number of (related) definitions. In the present work, we employed the common, and perhaps uncontroversial, view that a necessary condition of automaticity is that the process should be fast and goal-independent (see Moors & Houwer, 2006, for a review). Thus, if perspective-taking is automatic, one should expect it to occur when participants are engaged in a secondary task (i.e., detecting a target), and when several seconds of scene viewing are not required for the effect to occur. We also aimed to test the perspective-taking theory using the most recent paradigm that has been employed in support of the theory, that is, the 'ambiguous number' paradigm.

We examined the perspective-taking theory via the use of eye movement measurement. Eye movements provide a relatively non-intrusive online measure of attention which allows attentional mechanisms to be studied under more naturalistic conditions than many other visual cognition paradigms (Findlay & Gilchrist, 2003). Several studies have shown that eye movements are influenced by social cues and illustrate how people generally look at objects that are looked at by others (Fletcher-Watson, Leekam, Benson, Frank, & Findlay, 2008; Kuhn, Tatler, & Cole, 2009; Leekam, Hunnisett, & Moore, 1998; Zwickel & Vo, 2010). Furthermore, the overall time spent inspecting an object, in addition to the time taken to fixate the object, provides a valuable index of attentional allocation. Although overt gaze following has

been central to our understanding of ToM in infants (Butler, Caron, & Brooks, 2000; Caron, Kiel, Dayton, & Butler, 2002), eye movement measures have not been typically used to investigate the automatic perspective-taking claim (see Ferguson, Apperly, & Cane, 2017).

In the present Experiment 1, participants freely viewed everyday scenes that contained a model who either looked towards an object/area of interest or looked elsewhere. Orthogonally to the gaze direction, we manipulated whether the model's view of the object was occluded (by a natural barrier) or in full view. We predicted that participants would be faster to fixate the object when the person depicted in the scene looked towards it, i.e., a basic gaze cueing effect. Crucially, we predicted that this social facilitation *will* be modulated by whether the target object is visible or not to the model. Although some studies challenge the notion that mental state attributions influence gaze following (see above), this prediction was based on the fact that freely viewing a scene allows observers to employ higher mechanisms concerned with ToM. It is precisely under these circumstances that mental state attributions should influence gaze following. That is, when participants have time to consider what the model is looking at, i.e., non-automatically. In Experiment 2, participants viewed the same scenes but, rather than freely viewing the images, they performed a standard target discrimination task in which the targets were positioned at the object/area of interest employed in Experiment 1. In this scenario, we reasoned that participants' attentional set would be concerned with rapidly finding a prespecified target, thus vastly reducing the likelihood that they would consider what the model is viewing. We also used the barrier technique (Experiment 3) to examine the automatic perspective-taking claim and employed the centrally located gaze cue method. In the final experiment (Experiment 4), we used the

'ambiguous number' paradigm, together with the visibility manipulation, as a relatively new test of the perspective-taking theory.

Experiment 1

Method

Participants. Sixty (47 female) Brunel University students took part. Age ranged from 18 to 44 years (M = 20; SD = 5). All participants reported normal or corrected to normal vision.

Stimuli and apparatus. Twelve different scenes were photographed. Each contained one model who either looked towards an object/area of interest (e.g., a cup of coffee) or away from it. Each scene also contained an object that was located such that it could act as a barrier between the model and object (see Figure 1). Thus, there

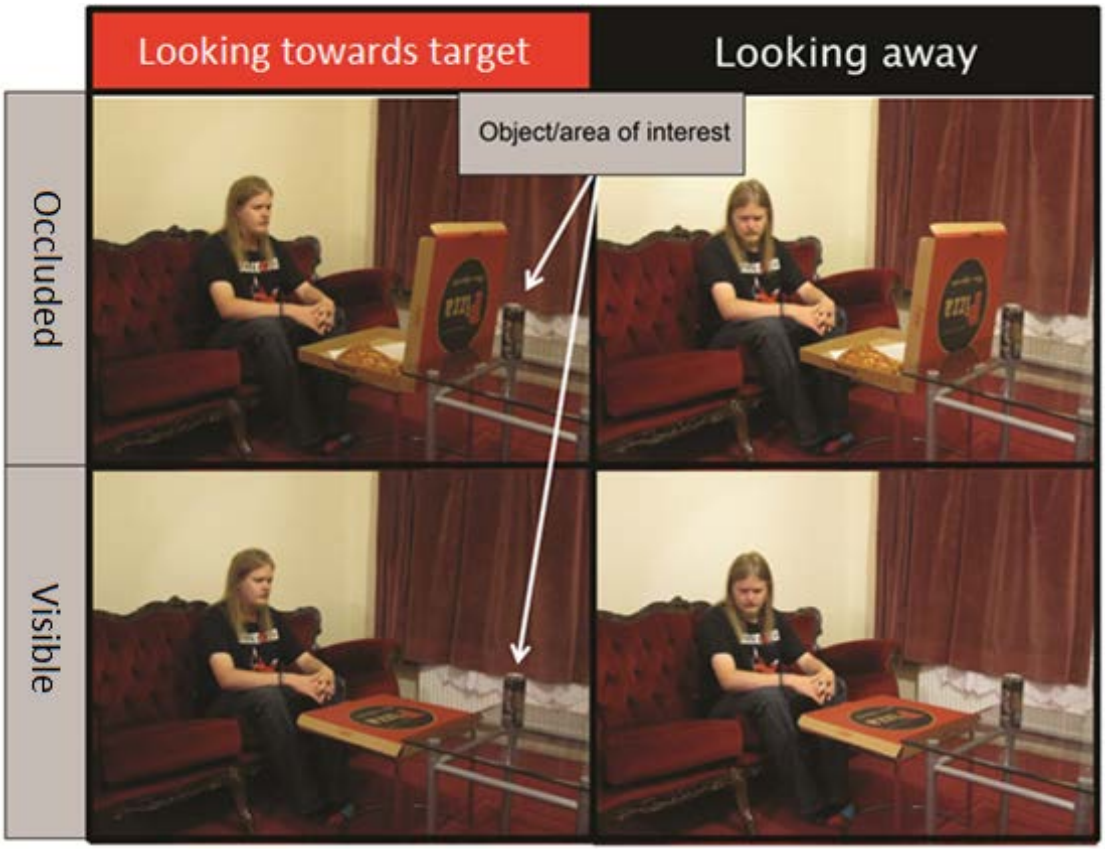

Figure 1: One of the image sets in which the model either looked towards the object of interest (i.e., drinks can) or away from it. The object was either visible to the model or not depending on the position of a barrier.

was a total of 48 images generated, i.e., 12 scenes, each with a valid and invalid gaze cue, and each with the object of interest being visible or occluded. Eye movements were recorded with a head-mounted, video-based eye tracker (EyeLink 1000; SR Research Ltd., Osgoode, Ontario, Canada), using a sampling rate of 1000Hz. They were recorded monocularly and analyzed using Eyelink Data Viewer (SR-Research). The images were presented on a 21-in CRT monitor (1024 x 768; 85HZ) using Experiment Builder presentation software (SR-Research), with a viewing distance of approximately 57 cm.

Humans are known to be sensitive to precise directions of gaze (Symons, Lee, Cedrone, & Nishimura, 2004). We therefore tested whether any absence of a cueing effect in the barrier condition of Experiment 1 would be because the model in those scenes was not able to look directly in line with the target object (because the occluder blocked its view). In our test, 18 participants (not used in Experiment 1) were presented with the 12 images showing the agent looking towards the target object with the occluding barrier. They were asked "if the x [e.g., pizza box] was not there, what would that person be looking at"? Overall, participants correctly identified the target object on 92.1% of trials. We then undertook a further variant of this test in which the occluding barrier was blacked out in all 12 scenes and 18 (new) participants were asked "What is that person looking at?" Correct identification rate was 94%. These two assessments of our stimuli thus demonstrate that there is little ambiguity concerning where the model was looking when the target objects were occluded by the barriers.

Design and procedure. We employed a 2 x 2 design with gaze (towards or away from object/area of interest) and visibility (visible, occluded) as within-participant factors. A Latin Square design was used (four lists) to ensure that each scene was only presented once. Each list contained 12 unique scenes, and an equal number of trials for each factor combination. The eye tracker was calibrated using a 9-point calibration and

validation procedure, after which participants were asked to freely view the scenes each for 5 seconds. That is, participants were simply asked to look at the pictures, without any further instructions. The images were presented in a random order. Each trial was preceded by a validation procedure consisting of a fixation point presented in the centre of the display.

Results and discussion.

For each image we calculated two different eye movement measures: 1) Time to fixate the object (time elapsing between the onset of the display and the first fixation on the interest area), and 2) Proportion of time spent fixating it (i.e., 'dwell time').

Time to Fixate

Figure 2 (left panel) shows mean RT for participants to fixate the object. Participants failed to fixate the target object on 28% of trials[1] and 12 participants did not look at the object in at least one of the conditions, which meant that the sample for this analysis was 48, rather than 60 participants.

A 2 X 2 anova with gaze (towards, away) and visibility (visible, occluded) as within-participant factors revealed a significant main effect of gaze, $F(1, 47) = 7.27$, $p = .01$, $\eta^2 = .13$, but no significant main effect of visibility, $F(1, 47) = .01$, $p = .075$, $\eta^2 = .002$. There was a significant gaze by visibility interaction, $F(1, 47) = 5.2$, $p = .027$, $\eta^2 = .10$. Post-hoc analyses showed that when the object was visible to the model, participants were significantly faster to fixate it when the model gazed towards it, as opposed to when the model looked elsewhere, $t(47) = 3.35$, $p = .002$. No such significant difference occurred when the object was not visible, $t(47) = .37$, $p = .71$.

[1] Participants were significantly less likely to ever fixate the target object when the person looked away from it than when the person looked towards it, but this difference was only significant in the visible condition (Wilcoxon test , p = .005) and not in the non-visible condition (Wilcoxon test, p = .15.

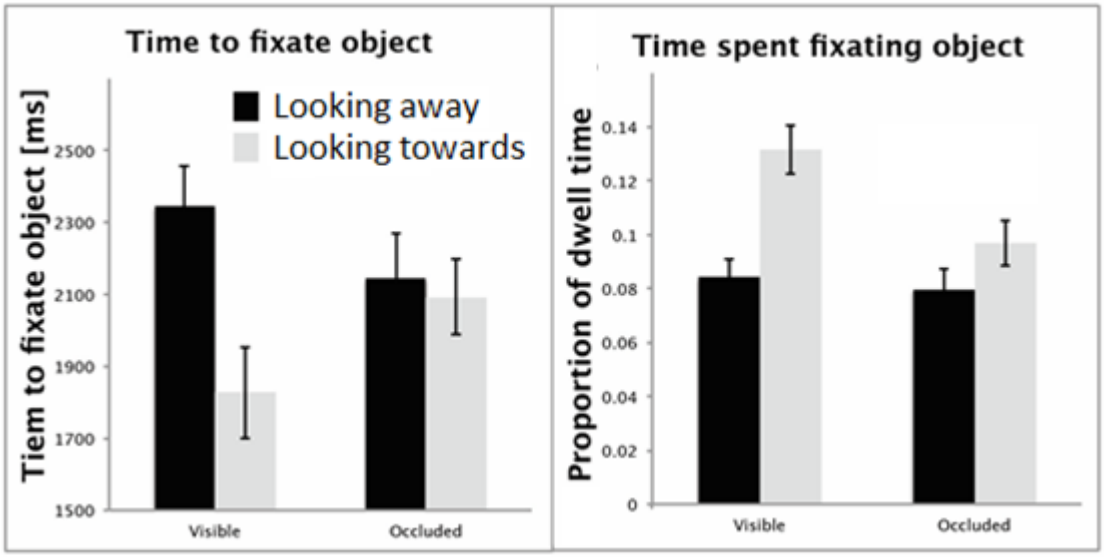

Figure 2. Time to fixate the object/area of interest (left panel) and time spent fixating object/area of interest (right panel) as a function of gaze direction and whether the object was visible or occluded. Standard error bars are also included.

Proportion dwell time

Figure 2 (right panel) also shows the time spent looking at the object. A 2 X 2 anova with gaze (towards, away) and visibility (visible, occluded) found a significant main effect of gaze, $F(1, 59) = 23.3$, $p < .0001$, $\eta^2 = .28$, and a significant main effect of visibility, $F(1, 59) = 9.58$, $p = .003$, $\eta^2 = .14$. There was also a significant gaze by visibility interaction, $F(1, 59) = 4.53$, $p = .038$, $\eta^2 = .071$. Post-hoc analyses revealed that participants spent more time fixating the object when it was looked towards in both the visible condition, $t(59) = 4.25$, $p = < .00005$, and occluded condition, $t(59) = 2.06$, $p = .044$.

One further analysis examined the time spent fixating the barrier. When it occluded the target, participants' spent significantly more time fixating the barrier when the model looked towards the target (proportion M = .21, SD = .09) than when the model looked elsewhere (M = .16, SD = .078; t(59) = 3.20, p = .002). However, this difference was not significant when the barrier did not obstruct the model's line of sight (looked at target, M = .12, SD = .061; looked elsewhere, M = .13, SD = .081; t(59) =

.51, p = .61). Indeed, a significant gaze by visibility interaction, ($F(1, 59) = 7.66$, $p = .008$, $\eta^2 = .12$) illustrates that the model's gaze only affected the amount of time participants spent fixating the barrier when it was directly in the line of sight.

The central findings from Experiment 1 is that participants were significantly faster to fixate the object/area of interest if the model gazed towards it. That is, a basic gaze cueing effect was observed. This however only occurred when the model could see the object, not when its view was occluded. These data thus suggest that mental state attribution *can* influence gaze following. As we set out in the Introduction, this effect is likely to be due to participants, under the condition of free viewing, having time to consider the mental state of the model, i.e., non-automatically. One does however have to consider the possibility that computation of the model's perspective *did* occur automatically, in the first few milliseconds of scene presentation, but then participants consciously computed this perspective during the rest of the five seconds of free viewing. In Experiment 2 we explicitly examined whether the present perspective-taking effect is likely to be automatic.

Experiments 2

Participants viewed the same scenes as were presented in Experiment 1. However, rather than freely viewing the images, they were instructed to search for a prespecified target (i.e., a horizontal or vertical line), typical of a standard visual search task. The targets were positioned either at a location that the model looked towards (superimposed on the objects identified in Experiment 1) or elsewhere. As with Experiment 1, the target could be seen by the model or was occluded. In order to examine the time course of any effect, we added a stimulus onset asynchrony (SOA) manipulation, in which the target either appeared simultaneously with the onset of the scene, or 300 ms later. Although Samson et al. (2010) reported a 'perspective-taking'

effect at 0 ms SOA (i.e., dots/targets and avatar appeared simultaneously) we included the additional SOA as a more liberal test of the automaticity claim. Any effect of the barrier at 300 ms SOA would at least provide evidence that the effect is relatively rapid. Given that participants are unlikely to have time to consciously consider the model's mental state, we predicted that although gaze would modulate target discrimination this would not be influenced by whether the model could see the target or not.

Method

Participants. Sixty (38 female) Essex University students took part. Age ranged from 18 to 46 years (M = 20; SD = 5). All participants reported normal or corrected to normal vision.

Stimuli and apparatus. We used the same scenes that were presented in Experiment 1, and superimposed a small green line (1.1° in length) that was either horizontal or vertical onto the target object. This line acted as the target that participants were required to respond to. Eye movements were monitored using an Eyelink 1000 (SR-Research), and data analysis was conducted using Data Viewer.

Design and procedure. We employed a 2 x 2 x 2 design with gaze (towards target, away from target), visibility (visible, occluded), and SOA (0 ms, 300 ms) as within-participant factors. The experiment began with 12 practice trials, in which all scenes were presented, followed by 192 experimental trials presented in a different random order for each of the participants. The eye tracker was calibrated using a 9-point calibration and validation procedure, and each trial started with a central fixation cross, which participants were asked to fixate before a trial was initiated. Participants were required to search for the target line and indicate as quickly as possible whether it was horizontal or vertical, by pressing the 'z' or 'm' key on a standard keyboard. Since the trial finished as soon as the target was detected, dwell time data are not meaningful and

thus were not further analysed. Manual reaction times and time to fixate the target were calculated as the difference between the onset of the target display and the time at which participants pressed the key (Manual), or they first fixated the target (time to fixate target).

Results and discussion

Manual RT.

RTs longer than 3000 ms were treated as outliers and excluded from the formal analysis (M = .13%). Figure 3 shows mean RTs for discriminating the target. An anova with gaze (away, towards target), visibility (visible, occluded) and SOA (0 ms, 300 ms) as within-participant factors found a significant main effect of gaze, $F(1, 59) = 6.41$, $p = .014$, $\eta^2 = .098$, replicating a typical gaze cueing effect. There was a significant main effect of SOA, $F(1, 59) = 601$, $p < .0001$, $\eta^2 = .91$, in that targets were detected more rapidly in the 300 ms SOA condition. There was no significant main effect of visibility, $F(1, 59) = 1.37$, $p = .25$, $\eta^2 = .023$. Additionally, there was a significant SOA by visibility interaction, $F(1, 59) = 5.46$, $p = .023$, $\eta^2 = .085$. Crucially there was no significant gaze by visibility interaction, $F(1, 59) = 3.73$, $p = .058$, $\eta^2 = .059$, showing that the gaze cueing effect was independent of whether the target was visible or occluded. There was no significant SOA by gaze interaction, $F(1, 59) = .036$, $p = .85$, $\eta^2 = .001$, and no significant SOA by visibility by gaze interaction, $F(1, 59) = 3.39$, $p = .071$, $\eta^2 = .054$. With respect to errors, there were no significant main effects nor interactions, all Fs < 1.40, all ps > .23.

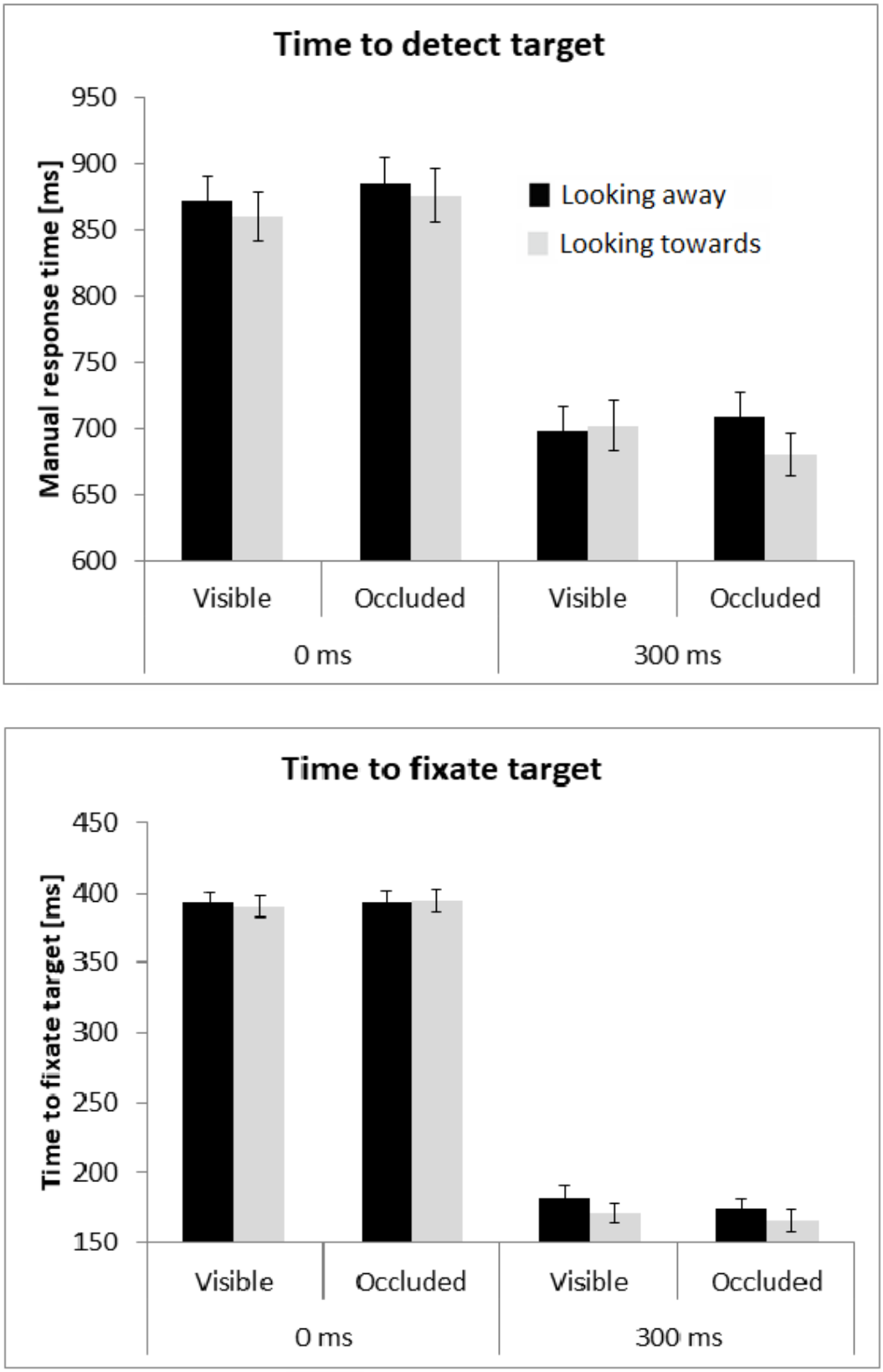

Figure 3: Mean manual RTs for detecting the target (upper panel) and time to fixate the target (lower panel) as a function of gaze direction, target visibility, and SOA. Standard error bars are also included.

Time to fixate target

Figure 3 also shows mean RTs to fixate the target. An anova, using the same factors and levels described above, found a significant main effect of gaze, $F(1, 59) = 5.14$, $p = .027$, $\eta^2 = .08$, again highlighting a general cueing effect. There was a significant main effect of SOA, $F(1, 59) = 2394$, $p < .0001$, $\eta^2 = .98$, but no significant main effect of visibility, $F(1, 59) = .44$, $p = .51$, $\eta^2 = .007$. Crucially, there was no significant gaze by visibility interaction, $F(1, 59) = .56$, $p = .46$, $\eta^2 = .009$, showing that the gaze cuing effect was not modulated by whether the targets was visible or occluded. There was no significant SOA by visibility interaction, $F(1, 59) = 3.33$, $p = .073$, $\eta^2 = .053$, and no significant SOA by gaze interaction, $F(1, 59) = 3.39$, $p = .071$, $\eta^2 = .054$. The three-way interaction was also not significant, $F(1, 59) = .011$, $p = .92$, $\eta^2 = 0$.

Overall, these results show a basic gaze cueing effect; targets were discriminated more rapidly when looked towards by the model. However, unlike Experiment 1, this effect was not modulated according to whether the model could see the target or not. This does not therefore support the automatic perspective-taking theory. Furthermore, an increase in SOA also failed to induce an effect of the barriers.

Experiment 3

Experiments 1 and 2 assessed the automatic perspective-taking theory using natural scenes and models that could appear in a variety of positions. Although automaticity is by most definitions efficient and rapid, it is possible that the attention system's 'race' to find the (salient) target in Experiment 2 was completed before the system had a chance to process what the gazing agent, located peripherally, could see. In Experiment 3 therefore, we employed the more typical central gaze cue (together with the barrier technique and saccade measurement) in which the agent always appears at fixation and looks to either the left or right (see Figure 4). A target then appears that

is either congruent with the gaze direction or incongruent. Presenting the agent centrally on each trial, rather than peripherally, should therefore have assisted in the computation of what the agent was looking at. Finally, in a further attempt to assist in the computation of any perspective-taking effect, we lengthened the shortest SOA (between scene and target presentation) to 80 ms. Thus, we presented targets at either 80 ms or 300 ms SOA.

Method

Participants. Thirty two (22 female) Essex University students took part. Age ranged from 18 to 44 years (M = 20; SD = 5). All participants reported normal or corrected to normal vision.

Stimuli and apparatus. The gaze cue agent was a female photographed sitting at a table, flanked by two balloons (Figure 4), and holding a physical barrier that obstructed her sight of one of the balloons. During each trial the two balloons changed

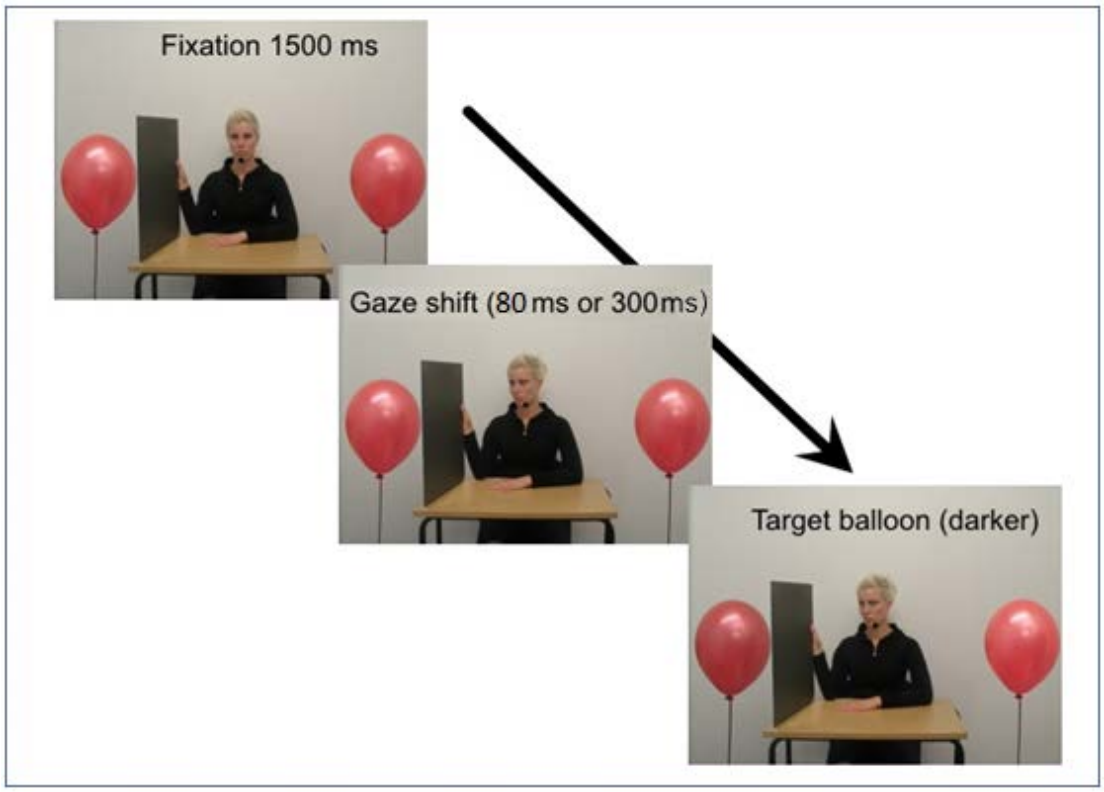

Figure 4. Example of an image presented in Experiment 3.

luminance. One of them became brighter and the other darker. The latter acted as the target. Target brightness and the fixation point were manipulated using Photoshop but all other aspects of the scene were present during photographing. Time to fixate the target was calculated as the time between the onset of the target display and the first fixation on the balloon.

Design and procedure. We again employed a 2 x 2 x 2 design with gaze (towards target, away from target), visibility (visible, occluded), and SOA (80 ms, 300 ms) as within-participant factors. The eye tracker was calibrated using a 9-point calibration and validation procedure, and each trial started with a central fixation point, which participants were asked to fixate. Each trial started with the agent looking straight ahead for 1500 ms after which she looked either towards the left or the right balloon. After 80 or 300 ms of this gaze/head shift (i.e., the SOA manipulation), participants were required to fixate the darker balloon as quickly as possible. They were then required to refixate and told to initiate the next trial with a button press. They were explicitly instructed to ignore the agent, and were also told that the barrier obstructed the woman's line of sight. Gaze and target location were independent of each other and presented randomly. The experiment was presented in two blocks (with presentation order counter balanced) and each block either had the barrier on the left or the right. Each block contained the same number of trials in which the actor's gaze was directed towards the target balloon or not, and whether the target was visible or occluded. The experiment was preceded by six practice trials followed by 128 experimental trials.

Results and discussion.

Data analysis was carried out using DataViewer (SR-Research). Each balloon was defined as a region of interest and we calculated the time to fixate the target balloon. Since the trial finished as soon as the target was detected, dwell time data were not meaningful and thus not further analysed. Data from 10% of the trials were excluded either because participants did not successfully fixate the target, or the equipment did not effectively track saccades.

Time to fixate target

Figure 5 shows mean RTs to fixate the target. An anova with gaze, visibility, and SOA as within-participant factors found a significant main effect of gaze, $F(1, 31) = 25.6$, $p < .001$, $\eta^2 = .45$, demonstrating the typical gaze cueing effect. There was also a significant main effect of SOA, $F(1, 31) = 20.6$, $p < .001$, $\eta^2 = .91$, but no significant main effect of visibility, $F(1, 31) = .98$, $p > .33$, $\eta^2 = .03$. There was no significant SOA by visibility interaction, $F(1, 31) = .005$, $p > .94$, $\eta^2 = .00$, and crucially no significant gaze by visibility interaction, $F(1, 31) = .16$, $p > .69$, $\eta^2 = .005$, showing that the gaze cueing effect was independent of whether the target was visible or occluded. There was no significant SOA by gaze interaction, $F(1, 31) = 1.8$, $p > .19$, $\eta^2 = .05$, and no significant SOA by visibility by gaze interaction, $F(1, 31) = .3$, $p > .58$, $\eta^2 = .01$.

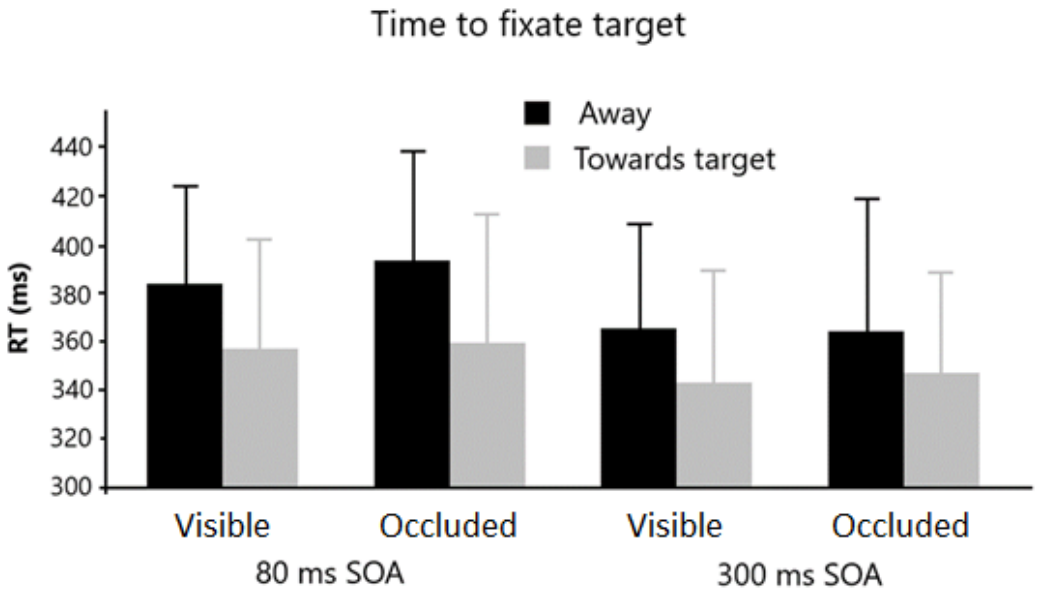

Figure 5: RTs and standard errors from Experiment 3.

Overall, these data show the gazing agent shifted observers' eyes in accordance with the direction in which she looked. Importantly however is the finding that this effect was not modulated according to whether the agent could see the target or not. As with the results of Experiment 2, this does not concur with the view that participants computed what the gazing agent could see.

Experiment 4.

As we set out in the Introduction, the notion that humans compute the visual perspective of other individuals has come primarily from the findings observed with the dot perspective and classic gaze cuing paradigms. A further 'perspective-taking' paradigm has been reported in which participants are required to identify a single number presented in a display (e.g., Surtees, Samson, & Apperly, 2016; Zhao, Cusimano, & Malle, 2015). As with the dot perspective and gaze cueing methods, an agent is present who looks towards the number. Importantly, the number is ambiguous such that it is different dependent on where it is viewed from. For example, in Figure 6

Figure 6. The ambiguous number paradigm. When the number ‘69’ is located on the table it appears as ‘69’ irrespective of whether it is viewed from the reader’s position or from the far side of the table, i.e., the model’s. When by contrast the number ‘68’ is presented it is seen as ‘89’ from the model’s perspective. Additionally, the left panel shows an example of the ‘visible’ condition employed in the present Experiment 4 and the right panel shows the ‘non-visible’ condition in which the agent is not looking at the number.

the number located on the table is the same from our position, as the viewer, and that of the agent sitting at the table. This is because ‘69’ appears the same when it is presented upside down. However, when the number ‘68’ is positioned on the table, it will appearas ‘89’ to the agent. Surtees, et al. (2016) found that RT to determine whether a number was ‘6’ or ‘9’ from the participant’s perspective was influenced by whether it appeared the same (‘consistent’) or different (‘inconsistent’) to an avatar. Specifically, RT was longer when the number was different to the avatar, e.g., ‘6’ to the participant and ‘9’ to the avatar. Although Surtees et al. did not argue that this particular effect was automatic, it does suggest that the avatar’s perspective was computed. A related procedure, also employing the ambiguous number method, was also reported by Zhao, et al. (2015). Rather than generating mean RTs from a large number of trials, Zhao et al. presented the kind of stimuli shown in Figure 6 and asked participants one question, “What number is on the table”. The authors found that approximately 40% of participants stated the number as seen from the avatar’s

perspective. As with Samson et al. (2010), Zhao et al. argued that participants had taken into consideration the avatar's perspective.

In the present experiment, we employed the Surtees et al. variant of the ambiguous number paradigm (i.e., mean RT over a number of trials) together with a variant of the visibility manipulation in which the avatar was either viewing the number on the table or was looking elsewhere. As with the dot perspective and gaze cueing paradigms, if the results obtained in the ambiguous number procedure are due to the avatar's perspective being assumed then no perspective-taking-like data should occur when the gazing agent is not looking at the number.

Method:

Participants. 33 (22 female) Essex University students took part. Age ranged from 18 to 46 years (M = 20; SD = 5). All participants reported normal or corrected to normal vision.

Stimuli and apparatus. A female was photographed sitting at a table in which she either looked down at a number physically located on the table (Figure 6, left panel) or looked to her side, i.e., away from the table and number towards the right hand wall of the room (Figure 6, right panel). One of four numbers was presented (from the participant's viewpoint); '68', '69', '88', and '89'. Unlike in Experiments 1-3, eye movements were not recorded.

Design and procedure. We employed a 2 x 2 design with number consistency (consistent, non-consistent) and number visibility (visible, non-visible) as within-participant factors. To reiterate these factors and levels, the consistent conditions were those in which the number would appear the same irrespective of whether viewed from one side of the table or the other (i.e., '88' and '69'). This contrasted the inconsistent conditions in which the number would be different (i.e., '68' and '89'). The visible

conditions were those in which the gazing agent looked directly at the number, and the non-visible conditions were those in which the agent looked away. Whereas the two visibility conditions were blocked, and presentation order counterbalanced, the consistency trials were presented randomly within block. Unlike our previous experiments, we blocked the visibility factor so that participants did not need to compute the agent's visual perspective trial-by-trial. This, we reasoned, would assist any perspective-taking effect.

Although it was not a factor of interest, we also blocked the mapping between consistency and hand response to ensure that any consistency effect would not be due to speed of response by one hand or the other. Thus, on half the trials the consistent condition corresponded with a left response whilst on the other half the consistent condition corresponded with a right response. This necessarily meant that the inconsistent condition was also distributed evenly across right and left responses. These consideration meant that four blocks of trials were presented to each participant; 1) Visible; left hand response when '68' appeared, right hand response when '69' appeared, 2) Visible; left hand response when '88' appeared, right hand response when '89' appeared, 3) Non-visible; left hand response when '68' appeared, right hand response when '69' appeared, 2) Non-visible; left hand response when '88' appeared, right hand response when '89' appeared. Each block presented 96 trials. Thus, there were 384 trials in total.

As with Surtees et al. (2016), we undertook a 'pre-test' experiment (N = 12) to ensure that RT to discriminate the two 'consistent' numbers that we would employ in the experiment proper (i.e., '69', '88') did not differ from the RT to discriminate the two 'inconsistent' numbers (i.e., '68', '89'). This pre-test was identical to the experiment proper with the sole exception that on each trial only the number was

presented and its immediate (beige) background. This was achieved by taking the images used in Experiment 4 and cutting away, via Microsoft Paint, all the surround. Results showed no significant difference in mean RT (consistent = 453 ms, SD = 55; inconsistent = 455 ms, SD = 54; t(11) = .38, p > .71).

Results and discussion

Figure 7 shows mean RTs for all four conditions. An anova with consistency and visibility as within-participant factors found a significant main effect of consistency $F(1, 32) = 17.4$, $p < .001$, $\eta^2 = .35$, but no significant main effect of visibility, $F(1, 32) = .54$, $p > .46$, $\eta^2 = .01$. There was no significant interaction, $F(1, 32) = 1.1$, $p > .29$, $\eta^2 = .03$. Using the same factors and levels as described for the RT data, no significant effects of errors was observed, all Fs < 1.2 and all ps > .27.

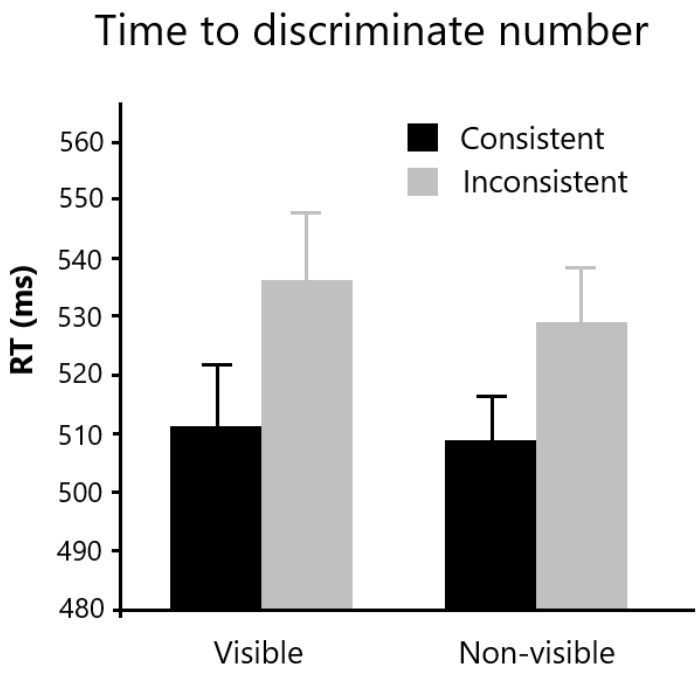

Figure 7. Mean RTs from Experiment 4. Standard error bars are also included.

These results replicate previous findings observed with the ambiguous number paradigm; RTs were longer when the number to be discriminated was different for the participant and avatar (i.e., inconsistent) relative to when they were the same (i.e.,

consistent). This finding on its own supports the notion that the avatar's perspective was computed. However, we have observed the same data pattern when the avatar was not looking at the target. If the avatar's perspective was indeed being taken no such effect should have occurred. This replicates previous findings showing that perspective-taking-like data can occur even when the gazing agent is not viewing the critical stimuli (e.g., Cole et al., 2015). Overall, the results from the present experiment do not support the perspective-taking theory.

General Discussion

The central aim of the present work was to evaluate the claim that a person's visual perspective is computed and that this occurs automatically. Following Cole et al. (2015), we manipulated what a gazing agent could see with the use of a physical barrier that sometimes occurred between a gazing agent and an object/area of interest. If the agent's visual perspective drives gaze-induced behaviour, it follows that no such gaze behaviour should occur when the agent cannot see the critical stimulus. Our results illustrate that when participants freely viewed images (Experiment 1), the model's gaze/head orientation influenced where people look. In line with previous research (Fletcher-Watson et al., 2008; Freebody & Kuhn, 2016; Zwickel & Vo, 2010) participants were significantly faster and spent significantly more time fixating objects that were looked towards by the model. However, this effect was not observed when a barrier interrupted the model's line of sight. This finding illustrates that under the conditions of free viewing, gaze following only occurred when the object located at the area of interest was visible to the actor. However, a different pattern of data occurred when participants were instructed to find a specific target, as opposed to freely view the scenes (Experiment 2). Although our participants were significantly faster to discriminate targets that were looked towards by the model (i.e., a gaze following

effect), this effect was not influenced by whether the model could see the target or not. We also found a facilitatory effect of a person's gaze direction when the target was occluded using the classic central (gaze) cueing paradigm in which the gazing agent was presented at the same location on every trial (Experiment 3). Finally, using the most recent paradigm to have been used in support of the perspective-taking theory (i.e., the ambiguous number procedure), we observed perspective-taking-like data even when the agent was not looking at the critical stimulus (Experiment 4).

It is our contention that during free viewing, participants have ample time to consider what a gazing agent can see. Indeed, in the present Experiment 1, our participants did not immediately fixate the area of interest; much of this time was spent fixating the model's face, allowing time to compute what the actor was looking at. When, by contrast, participants were instructed to perform a target discrimination task (i.e., Experiment 2), the target was rapidly fixated (< 500 ms), and often within one or two fixations, illustrating that the actor's gaze direction was processed within a relatively short time period. Moreover, participants rarely fixated the actor's face prior to fixating the target, suggesting that this gaze-induced attentional effect occurred even when gaze and head direction were processed peripherally (Hermens, Bindemann, & Burton, 2015). Thus, in the target discrimination scenario, rapid, or automatic-like, processing did not occur with respect to what the model could see. That is, the physical barriers did not modulate the gaze following effect.

As we stated in the Introduction, adult humans often follow others' gaze precisely because we would like to know what it is that another person is viewing. Thus, rather than suggesting that mental states do not modulate gaze following, we challenge the notion that this occurs automatically. Indeed, the present result can be seen as supporting much of what is known about gaze following and mental state attribution.

Reading other peoples' mental states is clearly an essential part of many successful social interactions, and an abundance of evidence supports the view that much of our behaviour is influenced by these metal states, including behaviour in young children. For instance, by 18 months, a child's gaze following can be modulated according to whether a critical object can be seen or not by the viewer (Butler et al., 2000; Caron et al., 2002).

Perhaps the most pertinent question associated with the issue of whether another person's viewpoint is computed automatically is what should we mean by automatic. In the current work, we employed the common notion that a necessary condition of automaticity is that the process should be fast and goal-independent. For instance, it should occur when participants have the goal of discriminating a simple target. The issue of automaticity became particularly important during the early days of the cognitive revolution in which research on selective attention was dominant. A number of theories made the distinction between preattentive and attentive processes, (e.g., Broadbent, 1958; Deutsch & Deutsch, 1963; Norman, 1968) with the former being viewed as occurring automatically (i.e., in 'parallel' and unlimited in capacity). Since then, a vast array of terms and concepts have been used to describe what is meant by automatic. In their extensive review of the issue, Moors and Houwer (2006) made the point that researchers have tended to associate automaticity with a (large) number of features including processes that are unintended, stimulus-driven, efficient, effortless, fast, unconscious, goal-independent, uncontrolled, autonomous, implicit, and difficult to alter or suppress. Of course, such features are not themselves always well-defined and are perhaps more indicative of the variety of language, than cognitive processes and mechanisms.

Although assessing automatic perspective-taking against *all* of these features is beyond the scope of the present article, some of the features are more easily addressed with respect to perspective-taking than others. For instance, when participants are asked to consider their own perspective, as was the case in the original version of the dot perspective paradigm, the altercentric intrusion that results is clearly not intended. By contrast, since our visibility manipulation only modulated gaze following in the free viewing condition the present results suggest that the perspective-taking process is not rapid. It is tempting to think that since computation of what an avatar can see is not relevant to the dot perspective task, any perspective-taking effect in that paradigm must be goal-independent. Indeed, Samson et al. (2010) stated that the avatar's perspective was computed spontaneously "even when it was not relevant to the task" (p1264). However, knowing when targets/stimuli are non-task relevant has itself been difficult to determine. Indeed, this is the central issue in the 'attentional control settings' debate (see for instance Theeuwes, 2004). In their seminal examination of how attention is distributed across a visual display, Folk, Remington, and Johnston (1992) argued that resource allocation is largely dependent upon an observer's goal-directed attentional set, or their attentional control settings. This was supported by Folk et al.'s initial work, and a subsequent large number of studies (e.g., Cole, Kuhn, Heywood, & Kentridge, 2009) showing that the propensity with which a characteristic of a stimulus is able to modulate attention is contingent on the stimulus sharing a feature that is relevant to an observer's task. One of the most important principles to come from this work is that task relevance can operate in extremely subtle ways. For instance, a stimulus can be comprised of and defined by many properties (e.g., colour, form, texture, luminance, contrast) and Folk and Remington (1998) showed that the likelihood of a stimulus modulating attention was largely influenced by the degree to which the stimulus

included properties shared with target. In other words, a seemingly 'task-irrelevant' stimulus can turn out to be task-relevant.

Another subtle influence on whether a task is truly goal-independent concerns the effects that other tasks and conditions have on a phenomenon within the same experiment. Such 'carry-over' effects reveal that responses on one trial can influence responses on other trials (e.g., Olivers, Humphreys, & Braithwaite, 2006; Tipper, 1985). Cole et al. (2016) pointed out that "*spontaneous* computation of others' perspective should not require observers to occasionally assume this perspective" (p. 166). In the basic dot perspective paradigm, participants are sometimes asked to consider the number of dots from the avatar's perspective whilst on other trials they are asked to do this from their own perspective. Any carry-over from the former task to the latter would negate the notion that the avatar's perspective is task irrelevant. Data from Samson et al. (2010) suggest that such carry-over does occur. When the perspective (i.e., the participant's or the avatar's) was manipulated within the same block (Experiment 1), the consistency effect was considerably larger (48 ms, $p < .01$) compared to when participants only performed the task from their own perspective (23 ms, $p < .05$; Experiment 3). One could of course argue that these data suggest that the carry-over effect only increases altercentric intrusion, rather than solely explaining the phenomenon. However, the fact that participants were instructed to consider perspective at all, if only their own, may have primed them to think about the avatar's perspective. If this was the case, the altercentric intrusion effect should not be considered as being automatic.

In sum, our results show that when given ample time to explore the visual environment, gaze following is modulated by another person's visual perspective. This altercentric intrusion did not occur when participants were required to rapidly

discriminate a target. That is, when participants were unlikely to be able to consider the person's visual perspective. This does not support the notion that anothers' visual perspective is automatically computed.

Raw Data:
All of the raw data are available from

https://figshare.com/s/77ea13abd9a7419009d7

References

Broadbent, D.E. (1958). *Perception and communication*. London: Pergamon Press.

Butler, S. C., Caron, A. J., & Brooks, R. (2000). Infant Understanding of the Referential Nature of Looking. *Journal of Cognition and Development, 1*(4), 359-377. doi: 10.1207/s15327647jcd0104_01

Caron, A. J., Kiel, E. J., Dayton, M., & Butler, S. C. (2002). Comprehension of the Referential Intent of Looking and Pointing Between 12 and 15 Months. *Journal of Cognition and Development, 3*(4), 445-464. doi: 10.1207/s15327647jcd3,4-04

Cole, G. G., Atkinson, M. A., Le, A. T. D., & Smith, D. T. (2016). Do humans spontaneously take the perspective of others? *Acta Psychologica, 164*, 165-168. doi: http://dx.doi.org/10.1016/j.actpsy.2016.01.007

Cole, G. G., Kuhn, G., Heywood, C. A., & Kentridge, R. W. (2009). The Prioritization of Feature Singletons in the Change Detection Paradigm. *Experimental Psychology, 56*(2), 134-146. doi: 10.1027/1618-3169.56.2.134

Cole, G. G., Smith, D. T., & Atkinson, M. A. (2015). Mental state attribution and the gaze cueing effect. *Attention Perception & Psychophysics, 77*(4), 1105-1115. doi: 10.3758/s13414-014-0780-6

Cole, G. G., & Atkinson, M., D'Souza, A., & Smith, D. (2017). Spontaneous perspective-taking in humans? Vision, 1, 17.

Deutsch, J. A., & Deutsch, D. (1963). Attention: Some Theoretical Considerations. *Psychological Review, 70*(1), 80-90. doi: 10.1037/h0039515

Driver, J., Davies, M., Kidd, P., Maxwell, M., Ricciardelli, P., & Baron-cohen, S. (1999). Gaze perception triggers reflective visiospatial orienting. *Visual Cognition, 6*(5), 509-540.

Driver, J., Davis, G., Ricciardelli, P., Kidd, P., Maxwell, E., & Baron-Cohen, S. (1999). Gaze perception triggers reflexive visuospatial orienting. *Visual Cognition, 6*(5), 509-540.

Findlay, J. M., & Gilchrist, . I. D. (2003). *Active Vision*. Oxford: Oxford University Press.

Fletcher-Watson, S., Leekam, S. R., Benson, V., Frank, M.C., & Findlay, J. M. (2008). Eye-movements reveal attention to social information in autism spectrum disorder. *Neuropsychologia, 47*, 248-257.

Folk, C. L., & Remington, R. (1998). Selectivity in distraction by irrelevant featural singletons: Evidence for two forms of attentional capture. *Journal of Experimental Psychology: Human Perception and Performance, 24*(3), 847-858. doi: 10.1037/0096-1523.24.3.847

Folk, C. L., Remington, R. W., & Johnston, J. C. (1992). Involuntary Covert Orienting Is Contingent on Attentional Control Settings. *Journal of Experimental Psychology-Human Perception and Performance, 18*(4), 1030-1044.

Freebody, S., & Kuhn, G. (2016). Own-age biases in adult's and children's joint attention. Biased face prioritisation, but not gaze following! *The Quarterly Journal of Experimental Psychology*, 1-19. doi: 10.1080/17470218.2016.1247899

Friesen, C. K., & Kingstone, A. (1998). The eyes have it! Reflexive orienting is triggered by nonpredictive gaze. *Psychonomic Bulletin & Review, 5*(3), 490-495.

Frischen, A., Bayliss, A. P., & Tipper, S. P. (2007). Gaze cueing of attention: Visual attention, social cognition, and individual differences. *Psychological Bulletin, 133*(4), 694-724.

Ferguson, H., Apperly, I., & Cane, J. (2017). Eye-tracking reveals the cost of switching between self and other perspectives in a visual perspective-taking task. Quarterly Journal Of Experimental Psychology, 70, 1647-1660.

Hermens, F., Bindemann, M., & Burton, M. A. (2015). Responding to social and symbolic extrafoveal cues: cue shape trumps biological relevance. *Psychological research*, 1-19. doi: 10.1007/s00426-015-0733-2

Kuhn, G., & Benson, V. (2007). The influence of eye-gaze and arrow pointing distractor cues on voluntary eye movements. *Perception & Psychophysics, 69*(6), 966-971. doi: 10.3758/bf03193934

Kuhn, G., Tatler, B. W., & Cole, G. G. (2009). You look where I look! Effect of gaze cues on overt and covert attention in misdirection. *Visual Cognition, 17*(6), 925-944.

Langton, S. R. H. (2018). I don't see it your way: The dot perspective task does not guage spontaneous perspective-taking, *Vision, 2*, 6.

Leekam, S. R., Hunnisett, E., & Moore, C. (1998). Targets and cues: Gaze following in children with autism. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 39*, 951-962.

Moors, A., & De Houwer, J. (2006). Automaticity: A Theoretical and Conceptual Analysis. *Psychological Bulletin, 132*(2), 297-326. doi: 10.1037/0033-2909.132.2.297

Morgan, E. J., Freeth, M., & Smith, D. T. (2018). Mental state attributions mediate the gaze cueing effect, *Vision, 2*, 11.

Norman, D. A. (1968). Toward a theory of memory and attention. *Psychological Review, 75*(6), 522-536. doi: 10.1037/h0026699

Olivers, C. N. L., Humphreys, G. W., & Braithwaite, J. J. (2006). The preview search task: Evidence for visual marking. *Visual Cognition, 14*(4-8), 716-735. doi: 10.1080/13506280500194188

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.

Samson, D., Apperly, I. A., Braithwaite, J. J., Andrews, B. J., & Scott, S. E. B. (2010). Seeing It Their Way: Evidence for Rapid and Involuntary Computation of What

Other People See. *Journal of Experimental Psychology-Human Perception and Performance, 36*(5), 1255-1266. doi: 10.1037/a0018729

Santiesteban, I., Catmur, C., Hopkins, S. C., Bird, G., & Heyes, C. (2014). Avatars and arrows: Implicit mentalizing or domain-general processing? *Journal of Experimental Psychology: Human Perception and Performance, 40*(3), 929-937. doi: 10.1037/a0035175

Surtees, A. D., & Apperly, I. A. (2012). Egocentrism and automatic perspective-taking in children and adults. Child Development, 83, 452-460.

Surtees, A., Samson, D. & Apperly, I. (2016). Unintentional perspective-taking calculates whether something is see, but not how it is seen. *Cognition, 148,* 97-105.

Theeuwes, J. (2004). Top-down search strategies cannot override attentional capture. Psychonomic Bulletin & Review, 11, 65–70.

Zhao, X., Cusimano, C., and Malle, B. F. (2015). In Search of Triggering Conditions for Spontaneous Visual Perspective-taking. In Noelle, D. C., Dale, R., Warlaumont, A. S., Yoshimi, J., Matlock, T., Jennings, C. D., & Maglio, P. P. (Eds.), Proceedings of the 37th Annual Meeting of the Cognitive Science Society (pp. 2811-2816). Austin, TX: Cognitive Science Society.

Symons L.A., Lee, K., Cedrone, C.C., & Nishimura, M. (2004). What are you looking at? Acuity for triadic gaze. Journal of General Psychology, 131, 451–469.

Teufel, C., Alexis, D. M., Clayton, N. S., & Davis, G. (2010). Mental-state attribution drives rapid, reflexive gaze following. *Attention Perception & Psychophysics, 72*(3), 695-705. doi: 72/3/695 [pii]

10.3758/APP.72.3.695

Teufel, C., Alexis, D. M., Todd, H., Lawrance-Owen, A. J., Clayton, N. S., & Davis, G. (2009). Social cognition modulates the sensory coding of observed gaze direction. *Current Biology, 19*(15), 1274-1277. doi: S0960-9822(09)01248-2 [pii]

10.1016/j.cub.2009.05.069

Teufel, C., Fletcher, P. C., & Davis, G. (2010). Seeing other minds: attributed mental states influence perception. *Trends Cogn Sci, 14*(8), 376-382. doi: 10.1016/j.tics.2010.05.005

Tipper, S. P. (1985). The negative priming effect: Inhibitory priming by ignored objects. *The Quarterly Journal of Experimental Psychology Section A, 37*(4), 571-590. doi: 10.1080/14640748508400920

Zwickel, J., & Vo, M. L. (2010). How the presence of persons biases eye movements. *Psychonomic Bulleting & Review, 17*(2), 257-262. doi: 17/2/257 [pii]

10.3758/PBR.17.2.257