Gaze allocation in a dynamic social situation: Effects of social status and speaking

Tom Foulsham*, Joey T. Cheng, Jessica L. Tracy, Joseph Henrich and Alan Kingstone

Department of Psychology, University of British Columbia

*Corresponding author: Department of Psychology, 2136 West Mall, Vancouver, B.C., Canada, V6T 1Z4. tfoulsham@psych.ubc.ca

Keywords: Eye movements, social attention, social status

Word count: 7620

Running header: Gaze in a dynamic social situation

Abstract

Human visual attention operates in a context that is complex, social and dynamic. To explore this, we recorded people taking part in a group decision-making task and then showed video clips of these situations to new participants while tracking their eye movements. Observers spent the majority of time looking at the people in the videos, and in particular at their eyes and faces. The social status of the people in the clips had been rated by their peers in the group task, and this status hierarchy strongly predicted where eye-tracker participants looked: high-status individuals were gazed at much more often, and for longer, than low-status individuals, even over short, 20-second videos. Fixation was temporally coupled to the person who was talking at any one time, but this did not account for the effect of social status on attention. These results are consistent with a gaze system that is attuned to the presence of other individuals, to their social status within a group, and to the information most useful for social interaction.

Introduction

Human environments have three defining characteristics that are often neglected by researchers investigating visual attention. First, they are very complex, requiring a gaze orienting system evolved to concentrate resources on the most informative objects at the expense of others. This system emerges as a natural consequence of the complexity of the environment and the existence of a foveated visual system: rather than perceiving everything in the visual field with equal fidelity, humans possess a central region of high-acuity which they shift to select items for more extensive processing. Thus, although attention research has traditionally been concerned with covert orienting to stimuli in simple arrays, investigations of attention in natural behaviour have relied increasingly on the measurement of eye movements (Findlay & Gilchrist, 2003; Hayhoe & Ballard, 2005). In particular, this field of inquiry seeks to identify the stimuli that are likely to attract eye fixations in different conditions. In some circumstances, these stimuli may be best described by their low level features—salient items such as a bright object on a dark background are particularly likely to be fixated (Foulsham & Underwood, 2007; Itti & Koch, 2000). However, in more realistic and complex situations, where people look is closely related to their actions, goals and cognitions in each environmental context (Ballard & Sprague, 2005; Land & Hayhoe, 2001; Yarbus, 1967).

A second defining characteristic is that, for humans, this environmental context tends to be social. More often than not, humans are immersed in an environment that includes other people, and a useful, and perhaps fundamental, goal of attention is to keep track of these individuals. Social attention allows people to monitor the behaviour, intentions and emotions of others, in order to guide their own actions, interactions, and learning processes. In laboratory studies, this phenomenon has been studied by showing that the faces, and in particular the eyes, of other people are salient items and powerful attentional cues. For example, schematic eyes direct attention reflectively in a manner thought to correspond to "gaze following" (Friesen & Kingstone, 1998). In images of complex natural scenes, viewers spend a large and disproportionate amount of time fixating other people, and in particular the eyes of others (Birmingham, Bischof, & Kingstone, 2008). Children and adults with autistic spectrum disorder, who show abnormal and reduced social interactions, may not look at people in scenes and movies to the same degree as normally functioning participants (Dalton, et al., 2005; Klin, et al., 2002), and these deficits in social attention may even be a causative factor in the disorder (Baron-Cohen, 1995).

Third, the natural environment is highly dynamic because the state, location and salience of the objects within it change over time. Many laboratory studies of visual attention are concerned with how people select items in space (for example targets in a search task) and the goals, stimuli and locations in these studies typically remain fixed (although some paradigms do require more dynamic attentional selection, .e.g. multiple object tracking, Pylyshyn & Storm, 1988; the attentional blink, Raymond, Shapiro, &

Arnell, 1992; task switching, Rogers & Monsell, 1995). The guidance of eye movements in natural scenes is often studied using static images (Foulsham & Underwood, 2008; Henderson, 2003), but it is not always clear how well this research transfers to the real world, where individuals and the visual environment are often moving, and where particular objects need to be fixated at certain times. In contrast, studies of gaze allocation in real world activities have typically emphasized the temporal patterning of eye movements in relation to action (Land & Hayhoe, 2001). For example, people look toward an object a few seconds before manipulating it, and they then move on to the next task in the sequence. Recently, some research has explored the distribution of attention and eye movements in movies, and these experiments have suggested that people show a relatively high degree of convergence in cognitive processing and the distribution of attention (Hasson, Nir, Levy, Fuhrmann, & Malach, 2004). In movies, gaze seems to be drawn to both low-level salient cues (such as sudden onsets and movement: Itti, 2005) and to semantic (whilst not necessarily salient) stimuli such as meaningful events and the actions of others (Klin, Jones, Schultz, Volkmar, & Cohen, 2002).

In this paper we investigate gaze allocation in a set of video clips showing three individuals conversing. Where and when do people look when naturally viewing such clips? While these are relatively controlled stimuli, they contain real people embedded in a realistic background and a dynamic situation, allowing an exploration of the spatiotemporal distribution of attention in a social context. Previous research would predict that the people in the clips will be potent at drawing the attention of observers,

even though there is no particular task requirement to fixate them. Which factors will determine who gets fixated, and when? The use of complex stimuli with several people adds a social dimension and permits us to investigate whether social psychological constructs have an effect on the allocation of eye movements.

One social factor that may be critical is the social status of the different individuals in the environment. In almost all social situations, humans readily develop hierarchically structured relationships, with some individuals exerting more influence on others and, consequently, attaining increased access to reproductively relevant resources (e.g., food, mates; Berger, Rosenholtz, & Zelditch, 1980). Indeed, individual differences in social status or rank may be ubiquitous in human social interactions (Boehm, 1993). Many other primates also form strong social hierarchies, and gaze following has been documented in several of these, such as monkeys (Emery, 2000). Ring-tailed lemurs also show spontaneous gaze following of other social group members in their natural environment, suggesting that social attention evolved early in species that interact in social groups (Shepherd & Platt, 2008). Chance (1967) hypothesized that social attention would reflect the dominance hierarchy of primate groups, such that the dominant individual receives the greatest number of glances, and a recent study of patas monkeys supported this prediction (McNelis & Boatright-Horowitz, 1998). It has also been demonstrated that the effectiveness of gaze as a social cue depends on the relative social status of the individual: low status monkeys reflexively follow the gaze of any familiar monkey, but high-status macaques will only respond in this way to other high-status animals (Shepherd, Deaner, & Platt, 2006).

In humans, observational studies have documented rank-biased attention among children, by coding their apparent gaze (Abramovitch, 1976; LaFreniere & Charlesworth, 1983; Vaughn & Waters, 1981). However, experimental evidence for effects of social status on attention in humans is scarce; similarly, very few studies have used eye-tracking methodology to assess the impact of status on humans' attention. One recent study reported that the social status of people depicted in an array of photographs influenced the extent to which these individuals attracted attention (Maner, DeWall, & Gailliot, 2008): the frequency of high-status males in an array was over-estimated, and an eye tracking study confirmed that people spent more time looking at men who were rated as high status. This is consistent with evolutionary theories positing that social status is important in mate selection, particularly for women choosing a male partner. However, consistent with evolutionary approaches predicting the importance of attention to high-status individuals for reasons other than mate choice (Henrich & Gil-White, 2001), high-status males were also potent in attracting the attention of male observers.

Although these findings suggest that the social status of targets in a display may influence the amount of attention they receive, they are also somewhat limited. Maner et al manipulated social status by editing photographs to show individuals wearing either professional or casual attire, and their stimuli were static photographs isolated on a blank screen with no social context, no movement, and a task that placed few demands on the attentional system. In contrast, here we measure gaze while observers watch video clips of a real social interaction, and social status is quantified on the basis

of previous ratings made by peers who participated in the interaction. If social status affects the distribution of gaze in this study, it will provide evidence i) that attention is guided, top-down, by social attributions rather than just by feature salience and ii) that social status plays a role in early human information processing.

Method

Participants

25 students participated in the experiment. All were recruited through the University of British Columbia Human Subject Pool, and they gave their full informed consent and received course credit in return for participating. All participants had normal vision and did not wear glasses. After the experiment, it was confirmed that the participants were unfamiliar with the people they viewed in the experimental video clips.

Stimuli and design

The experimental stimuli consisted of four sets of video clips. Each set was derived from a previous experiment (Cheng, Tracy, Henrich, Foulsham, & Kingstone, in prep) in which groups of unacquainted undergraduates completed an interactive decision-making task while being recorded by an unconcealed high-definition video camera with built-in microphone positioned in front of them. The decision-making task concerned a

hypothetical situation requiring participants to rank a list of items for their use in a survival situation (i.e. “which items would your group need to survive if marooned on the moon?”). Participants were given 30 minutes to discuss this task in groups of 6, sitting around a table with three people on each side, before deciding on a group answer. To incentivize correct responses, participants knew that if the group’s final response was close to the correct answer, each participant would be given a monetary bonus. The videos used in the present research featured the three individuals on one side of the table. Figure 1 depicts a schematic of the scene and the layout of the resulting video frames. Four representative videos were chosen for the eye tracking study. In each case, the three individuals in the video (hereafter, the “targets”) were classified according to social status scores from the original group-interaction experiment. Specifically, in that previous study, after the group task all group members rated the social status and influence of each target, among a battery of other judgments (3 items on a 7-point scale, e.g., “this person led the task”). Ratings were made in a round-robin fashion then aggregated across peers. The four sets of clips used were chosen because peer-rated scores revealed clear relative status differences of the targets within them; on average there was a 2.5 point difference in mean status ratings (overall SD=1.4) between two of the targets, with the third falling in between, suggesting that these individuals could be considered high, low and medium status. Given these differences, in subsequent analyses we were able to compare the degree to which people paid attention to targets of each status level, by taking the mean across the high, medium, and low status targets in the four videos. The mean (and standard

deviation) peer-ratings for each type of target (on our 7-point scale) were 5.78 (0.6), 4.99 (0.2) and 3.25 (1.5) for high, medium and low status respectively.

In this experiment, we were particularly interested in whether social status made a difference to an observer's gaze allocation, even when the observer had only brief

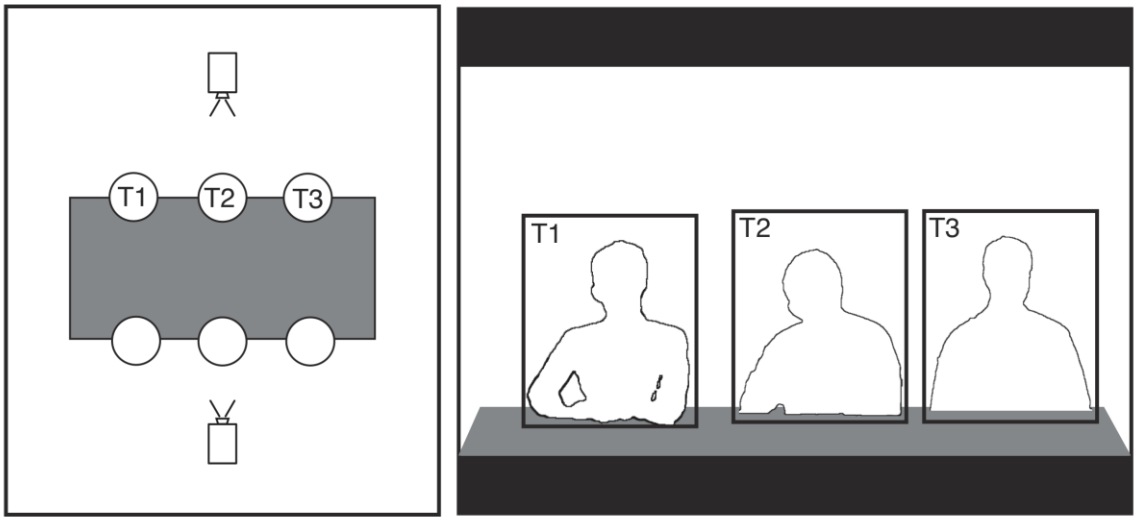

Figure 1. Stimuli production and layout. The videos were filmed using a camera facing each side of a table and capturing three of the people in the group (e.g., targets T1-T3; left panel). Frames from the clip featured these three people sitting side by side at a table (right panel, illustrating approximate size of the targets). The ROIs used to define fixations on the targets are shown as black boxes.

exposure to the target individuals. Given the difficulty of analyzing eye movements in video, and of maintaining an accurate track over long periods of time, we used 6 twenty-second clips for each set of targets. One research assistant blind to the study's hypotheses was instructed to choose 6 clips from each video that featured group

members negotiating and reaching a pivotal decision, so as to capture those moments where status dynamics were most salient.

The clips were cropped and formatted as digital movie files with dimensions of 1024 by 768 pixels and a frame rate of 30 fps. The aspect ratio of the original clips was 16:9, and thus a black border was added above and below the video image. The Xvid video codec (www.xvid.org) was used as it offered superior playback, as well as extremely accurate timing, which meant that the eye tracking apparatus could log exactly which frame was on the screen at any one time. Sound was played via an ASIO sound card, which maintained synchrony between video and audio. Each participant saw all six clips from one set in a random order. The set of clips seen by each participant was determined randomly, and each set of clips was seen by 6 participants, with the exception of one set that was seen by 7 participants.

Apparatus

The videos were shown on a 19-inch colour monitor with a refresh rate of 60 Hz. Participants used a headrest, which minimized head movements and ensured a constant viewing distance of 60cm, which resulted in an effective screen size of 40˚ by 31˚ of visual angle. At this distance, the visible area of the video frame was approximately 40˚ by 23˚. Sound was played through a pair of speakers positioned on either side of the monitor.

Eye movements were recorded using the Eyelink II system, which uses a head mounted camera. Pupil position was recorded monocularly from the video image of the right eye at 500 Hz. The Eyelink system used an on-line parser to extract fixations and saccades from the eye position samples, using velocity (30˚/s) and acceleration (8000˚/s^2) thresholds.

Procedure

The experiment began with the instruction that the participant should watch the clips as if they were in the room with the targets. More specifically, they were instructed to "imagine that you're in the room with these people, working on the task. Please think about which of the people in the group you would want to work with in a subsequent task". The sound volume was adjusted for each participant, and the eye tracker was calibrated with a 9-dot calibration routine that presented dots one at a time in known locations on the screen.

The trials then began. In each of the six trials, a drift-correct marker was first presented in the centre of the screen, and participants were required to look at the dot and press a key on the keyboard when central fixation was attained. This had the effect of constraining the initial fixation position to the centre of the screen, and correcting the eye tracker for any eye drift. The clip then appeared and the video and audio were played at normal speed for their duration of 20 seconds. Eye movements during this time were recorded, along with a record of timestamps indicating the onset time of each frame of the video. All 6 trials proceeded in this fashion.

Analysis and results

General viewing behaviour

We first assessed how participants responded to the clips by looking at the general eye movements they made.

Participants made an average of 49 fixations (SD=8.4) during each 20s clip, with fixations having a mean duration of 377ms (SD=83). The saccades between these fixations had a mean amplitude of 6.6˚ (SD=1.2). In all subsequent analysis, the fixation at clip onset was not included, because its central position was constrained by the procedure preceding the onset of the clip. To move beyond these simple descriptives, we quantified the attention given to the targets in the clip by defining a region of interest (ROI) around each person. This region was a rectangle with dimensions 10.9˚ by 14.1˚, a size that was kept constant for all targets. In most cases, there was relatively little movement of the targets within a clip, but for this first analysis the ROIs were large enough to encompass the targets throughout the whole clip. The ROIs for one clip are depicted in Figure 1 (right). Using these ROIs, we classified fixations as landing on one of the targets or on the background of walls, furniture and blank screen.

Across all clips and observers, an average of 77% of all fixations landed on the targets. It was relatively rare for the observers to look at the empty and static furniture and background. The ROIs covered 37% of the screen area, so if fixations were

uniformly distributed we should expect approximately this proportion of fixations to land on the targets. The fact that many more fixations were spent looking at the targets in the clips than this mean chance expectancy is preliminary evidence that participants focused their attention on the targets. A possible problem with this interpretation is that fixation distributions in a range of stimuli tend to be highly centralized (Foulsham & Underwood, 2008). As one of our ROIs was central, close to where viewing began and where participants tend to fixate, it might be that this underlies the tendency to fixate the targets. However, this explanation is unlikely to account for the data: peripheral targets were also fixated much more often than we would expect given their area, despite being further from the centre of the screen (44% of fixations landed on the left or right target, which together covered just 25% of the screen area). Thus, the people in the clips were potent at attracting fixation. Our subsequent analyses examined whether this varied as a function of these targets' social status.

Gaze allocation and social status

Each clip had three targets, classified as high, medium or low social status. We analyzed the eye movement data using repeated measures ANOVA with one within-subject factor of social status. Table 1 shows the measures taken for each level. First we considered the proportion of fixations that landed on the different targets.

	Target social status		
	High	**Medium**	**Low**
Mean proportion of fixations	0.35 (0.02)	0.28 (0.01)	0.14 (0.01)
Total fixation duration per clip (s)	6.47 (0.47)	4.86 (0.33)	2.30 (0.18)
Mean gaze duration (ms)	994 (74)	767 (68)	669 (45)

Table 1. Means (with standard errors in parentheses) for the different measures taken, as a function of social status.

Status had a reliable effect on the proportion of fixations on the target, $F(2,48)=31.7$, $p<.001$. There were more fixations on high-status targets than on medium-status targets, who received more fixations that low-status targets (all planned comparisons $p<.001$). This difference was quite pronounced. For example, medium-status targets received twice as many fixations as low-status targets, and high-status targets received even more attention.

An alternative way to measure the amount of attention paid to the different individuals in a clip is to look at the fixation time committed to each target. This was defined as the sum duration of all the fixations on each target, and it was averaged across the six clips to give the total fixation time per 20s clip. This measure reflects differences in how long observers looked at one target on each occasion, over and above the number of fixations. As previously, there was an effect of social status, $F(2,48)=34.1$, $p<.001$. Pairwise comparisons showed the same pattern as the previous

analysis: observers spent the most time looking at the high-status target, followed by the medium-status target, with the low-status target being inspected for the least amount of time (all $p<.01$).

The measures so far demonstrate that social status had an effect on the amount of attention given to the people in the clips. These measures were taken across a whole 20s video, comprising 10-20 fixations with a total duration of several seconds. An alternative question concerns how long the targets were gazed at on each visit, before a different person was inspected. For example, it is possible that high status individuals are looked at more often, and also that they hold an observer's attention for longer on each occasion that they are looked at. To explore this, we measured the mean gaze duration. A gaze was defined as the sum duration of all consecutive fixations on a target, with each gaze ending with a shift to a new target or to the background. On average, gazes were 810 ms, which corresponds to 2 or 3 fixations before shifting to a different region. Mean gaze duration was affected by social status, $F(2,48)=12.9$, $p<001$. The average length of each gaze on the high-status person was reliably longer than that on either of the other targets ($p<.05$). The low-status person received the briefest gazes, although the comparison between medium and low status fell short of significance.

Although the effects of social status on fixation behaviour are interesting, it is important to rule out more basic factors. One such factor is the spatial position of the people in the clips. As previously mentioned, people tend to fixate close to the centre of an image or video, and although seating was assigned to targets on a random basis,

those seated in the center may have taken on a high-status role as a result of their position. In fact, the low-status target was never positioned in the centre. Thus, centrality could explain the attentional bias away from these targets. However, in three of the four groups the high-status target was positioned on one side or the other, making centrality unlikely to explain the advantage for high-status over medium-status individuals. To further explore this issue, we conducted an additional analysis, comparing medium- and high-status targets in different positions. For this analysis, and for all those that follow, we focused on the proportion of fixations allocated to the different types of target, as the results from this measure and that of total fixation duration were identical. Figure 2 shows the proportion of fixations for the different types of target, both when they were positioned in the centre, and when they were positioned at the sides. It is clear from the graph that, although central targets were more likely to be fixated, the effect of social status was very similar at both spatial positions. High-status targets received more fixations on average than medium-status targets, both when they were each on the side of the display, and when they were both in the centre (both $t(23)>2.6$, $p<.02$). This is good evidence that the effect of status is not just an artifact of spatial position. Data for the low-status individual in the centre was not available because this target was positioned at the side in all clips, but given the results for high and medium-status targets, the low status targets' position is unlikely to have substantially influenced our results. To summarize this analysis, although seating position did matter (presumably because of a bias for fixations on the centre of the display), social status had an impact on attention at all seating positions.

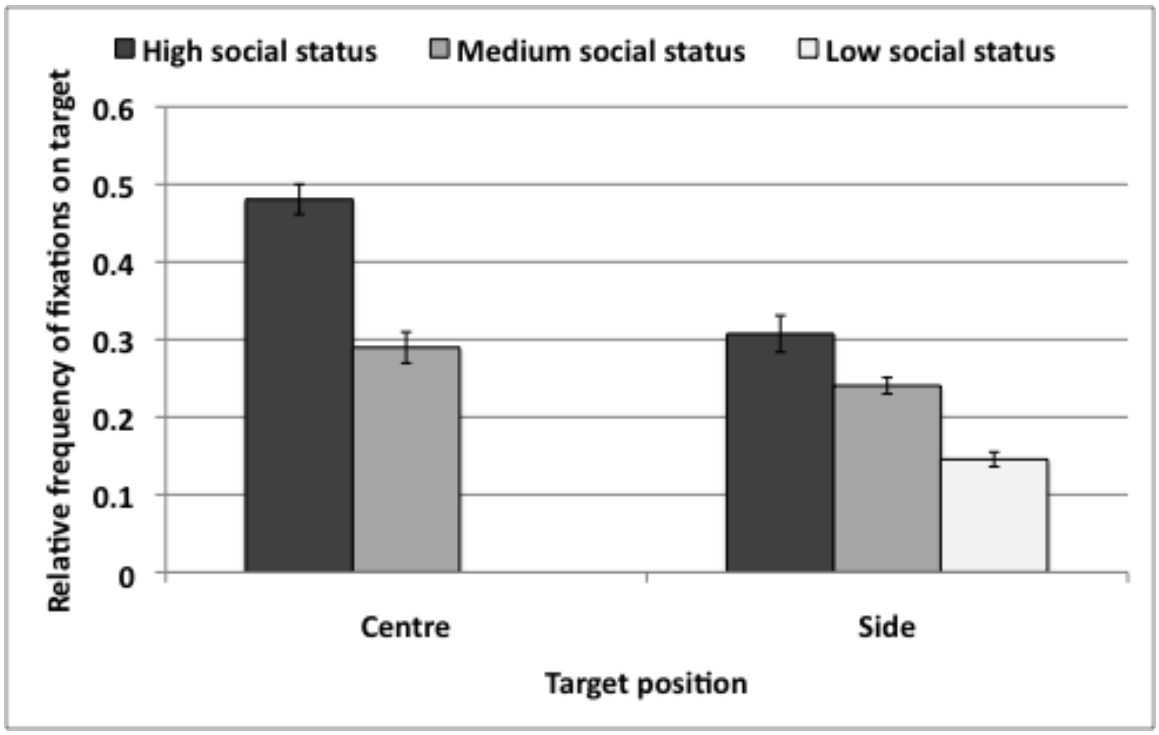

Figure 2. The proportion of fixations on targets that appeared at the centre or the sides of the group. Bars show the mean with standard error bars.

Gaze allocation and speaking

The eye movements of observers were sensitive to social status, and it is interesting to demonstrate this with complex stimuli and over only a short clip. What target behaviours underlie this effect? A strong candidate is the verbalizations of the individual. If high-status targets do most of the talking, and observers tend to look at the person speaking, this would explain our previous results. This would not be a trivial result, but it is important to ask whether status might have an effect in addition to that moderated by verbalizations.

Our eye movement methodology allowed us to look at the distribution of attention over time, with a high temporal resolution. To investigate how this distribution was related to verbalizations, we compared the fixation data to a record of who was talking at each moment in the clips. This record came from a trained independent observer, who watched all the clips and logged the beginning and end of each utterance. Specifically, we used custom-designed software that played the clips at a slow speed and allowed the observer to press one of two keys to indicate that a target had started or finished talking. This was repeated three times per clip (once for each target), and the result was a frame-by-frame timing matrix that showed which people, if any, were talking at any time (see Figure 3, top). As one might expect, the amount of time a target spent talking was related to their social status (one-way ANOVA across clips, $F(2,71)=11.4$, $p<.001$). High-status individuals spent the greatest proportion of the clips talking, followed by the medium-status targets and the low-status targets (means across all of the clips=26%, 19% and 5% respectively). Pairwise comparisons demonstrated that the low-status targets spoke for reliably less time per clip than either the high-status or the medium-status targets (both $p<.005$). The difference between high-status and medium-status targets was not significant. Importantly, the absence of a significant difference in speaking time between high- and medium-status targets suggests that the reported attentional differences between these targets cannot be solely explained by speaking time.

To control for both position and speaking time directly, we ran an analysis by target, comparing the average proportion of fixations that each target received in each

clip but adding target position (centre or side) and the proportion of time this target spent talking (in this clip) as covariates. This ANCOVA procedure statistically adjusted the dependant variable (mean proportion of fixations) to partial out the effects of speaking time and position. As expected from our previous analyses, both seating position ($F(1,67)=12.3$, $p<.005$) and talking time ($F(1,67)=35.9$, $p<.001$) influenced the attention given to each target. Targets who sat in the centre and spent more time talking were fixated most often. Most important, however, social status continued to have a reliable effect on the allocation of fixations over and above that predicted by the seating position and speaking time of the target ($F(2,67)=16.8$, $p<.001$). The same hierarchy of attention was seen, with high-status targets being fixated more often than medium-status targets and low status targets receiving the fewest fixations (all pairwise comparisons $p<.05$).

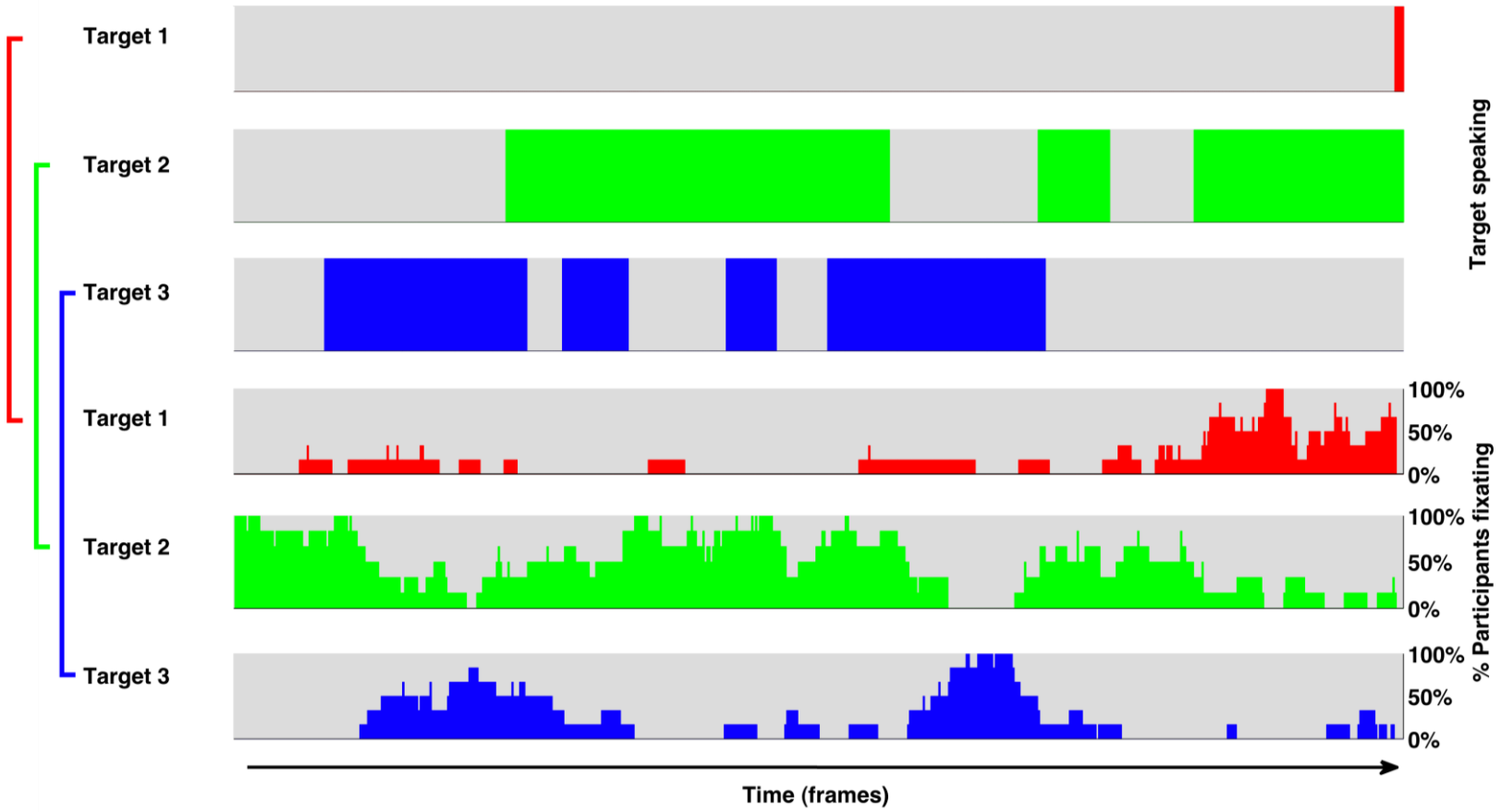

Figure 3. The synchrony between gaze and talking for one example clip, with time along the x-axis for a duration of 20s. The top three panels show whether each of the three targets in the scene (numbered 1 to 3 from left to right) was speaking at each point in time, with a solid bar indicating that they were. The bottom three bars show the proportion of observers watching the clip who fixated each of these people over the same time course. In many cases, a peak in participants looking at an individual coincides with that individual talking. In this clip, target 1 was low social-status, target 2 was high-social status and target 3 was medium-social status.

Figure 3 shows a graphical representation of one of the clips from the experiment. This visualization compares the record of who was speaking at any point in the clip (top three panels), to the proportion of observers who were fixating each target at that time (bottom three panels). At several points in this figure there is a tendency for observers to fixate the person who is talking. To explore this further we categorized all the fixations in a clip according to who was talking in the frame at which the fixation started. Because we are most interested in eye guidance, it is appropriate to categorize fixations with regard to their start time, as this will reflect the aspects of the target that attracted gaze toward them, rather than changes that occurred while the observer was fixating. This allowed a comparison between the proportion of fixations that were directed at the person talking and those that were on another target or on the scene background. Table 2 summarizes the relationship between who was speaking and who was being fixated.

As found in previous analyses, in general, when a target was talking participants were most likely to look at that person, and this can be seen in the relatively high values along the diagonal in Table 2. Did this trend vary according to the status of the person speaking? When the high-status person was talking, target status had a reliable effect ($F(2,48)=74.1$, $p<.001$). In this case the high-status speaker was fixated on almost half of all fixations, but on those occasions when someone else was fixated while the high-status person was talking, it was more likely to be the medium-status target than the low-status individual (all levels different at $p<.001$).

		Target speaking				
		High-status	**Medium-status**	**Low-status**		**Nobody**
Target being fixated	**High-status**	49%	22%	30%		34%
	Medium-status	22%	47%	23%		26%
	Low-status	11%	11%	31%		16%
	Background	18%	20%	16%		24%
	Total	100%	100%	100%		100%

Table 2. The relative frequency of fixations on each type of target, and on the background, expressed as a proportion of the total made while each target was speaking. Each cell shows the mean across participants, taking into account the differences in how often each target spoke. The first column, for example, shows who was fixated during the time that the high-status target was speaking.

The targets also received different amounts of attention when the medium-status target was speaking ($F(2,48)=43.3$, $p<.001$), with the person who was talking (in this case the medium-status target) again receiving the most fixations. However, when fixations were not on the medium-status target, the high-status target was more likely to be fixated than the low-status target ($p<.01$), even though neither of these targets were speaking. The low-status person was the least potent at attracting fixations when he/she was talking, and on these occasions participants were almost as likely to look at the high-status target as the speaker. There was no effect of status when the low-status target was talking ($F(2,48)<1$) and none of the pairwise comparisons were different. The

clearest demonstration that the effect of social status on gaze can be dissociated from speaking is apparent from the pattern of results on occasions when nobody was speaking: looking only at these fixations, there was an effect of social status, $F(2,48)=16.3$, $p<.001$, showing precisely the same pattern as observed previously: the high-status target was fixated more than the medium-status target, with the low-status target receiving the least attention (all comparisons $p<.05$). Thus, although people tended to look at the person speaking, social status remained important even when nobody was talking.

An alternative way of analyzing the fit between gaze and speaking is to use cross correlation. This technique analyzes the correlation between two signals over time, and it provides a correlation coefficient when the two signals are perfectly aligned (the "zero lag"), as well as when one signal is shifted relative to the other. In our case, we computed a cross correlation for each target, in each clip, between the record of speaking and the proportion of observers watching that clip who were fixating that target. We can then ask a) whether this correlation over time is statistically different from zero, and b) whether the highest correlation occurs at the zero lag. If the highest correlation were found at a different lag, it would suggest that there was a temporal delay between gaze and speaking. For example, observers might have looked at individuals a few frames after they started speaking. To give an estimate of the correlation we would expect by chance, we also made two sets of control comparisons. First, we compared the fixation record from each target and clip to the speaking record for all other targets and clips, which gives a baseline similarity between any two random

gaze and speaking signals. Second, we compared the gaze data from each target and clip with the speaking record of the same target in each of the 5 other clips in which that target appeared. This "matched target" comparison gives a measure of the chance correlation expected between fixations on a person and the speech of that same person in other situations. Table 3 shows the results of these analyses.

	Observed data	**Random control data**	**Matched target control data**
Cross correlation at zero lag	0.38 (0.03)	-0.01 (0.005)	-0.01 (0.01)
Maximum cross correlation	0.45 (0.03)	0.12 (0.004)	0.13 (0.01)

Table 3. Summary statistics from a cross correlation analysis of speaking and fixation. Cells show the mean (and standard error) correlation across all clips and targets.

Several interesting points can be drawn from this analysis. First, the cross correlation between a person speaking and their being fixated was reliably greater than zero. In comparison, the control data sets of fixations matched to the speaking data from other clips produced no correlations at the zero lag and much smaller correlations

when maximally aligned. Second, this correlation was higher still if we assume that there is a temporal offset in the relationship between speaking and fixation. The average lag at which the highest correlation was found can show the direction of this offset. Across all comparisons, the median lag at which the highest correlation between speech and fixation was found was -5 frames. Surprisingly, the negative offset indicates that, on average, correlations were higher when fixations were compared with the speaking that was going to take place 5 frames in the future. In other words, people tended to look at the speaker slightly (~150ms) before they started talking. This pattern was seen for all three types of target.

Regions of interest analysis

The previous analyses were based on relatively large areas of interest covering the whole of each person within the clip, and they showed that the targets were very frequently inspected. Which part of these targets was most potent at drawing gazes? There is a large literature showing the importance of faces, and in particular eyes, in drawing attention (Kingstone, 2009). In static photographs, people often spend most of their time looking at the faces and eyes of the people in the scene (Birmingham, et al., 2008). We therefore looked to see if the same was true in our dynamic movie clips, and also if this varied with social status. Because our targets would have moved slightly over the 20s clips, we first needed to define moving regions of interest. This was done by hand using custom software in MATLAB. Each clip was played at a slow speed, and a

mouse cursor was moved to follow the region in question, resulting in a record of where that region was at any frame in the movie. We did this for both the head region (which was kept to a standard size of 3.9˚ by 5.8˚) and the eye region (3.9˚ by 1.9˚) and for each target person. Fixations could then be labeled according to their location in the frame at which the fixation started. For example, a fixation was classified as on the eyes if, on the frame where it started, its spatial coordinates lay within the eye region. Figure 4 shows an example of these regions, and the relative frequency of fixations on the eyes, the rest of the head (defined as head minus eyes) and the rest of the body (defined as the original target ROIs minus the head).

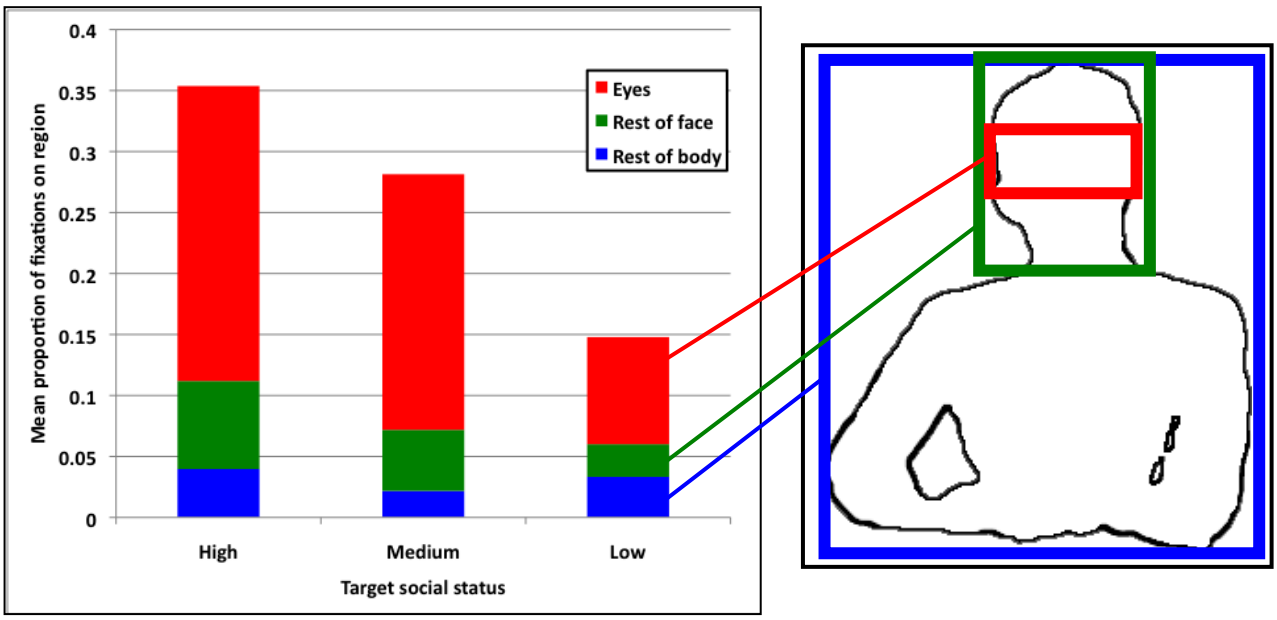

Figure 4. Measuring the amount of gaze given to different parts of the people was accomplished by defining moving areas of interest for the eyes and head (the relative sizes of which are depicted with a diagram of one target in the right panel). The proportion of fixations on each of these regions, averaged across all observers, is shown

in the left panel.

We analyzed the proportion of fixations on each region using repeated measures ANOVA with two factors: social status (high, medium, or low) and region of interest (eyes, rest of head, or rest of body). As previously, there was a significant effect of status, $F(2,48)=31.8$, $p<.001$. There was also a main effect of region of interest, $F(2,48)=74.7$, $p<.001$. Summing across all targets, the mean probability of a fixation landing on someone's eyes was 54%, much greater than gazes to the rest of the face (15%) or to the body (10%). All these averages were reliably different (all $ps<.05$). There was also a reliable interaction, $F(4,96)=22.5$, $p<.001$, showing that the potency of the different regions at drawing fixations varied with the social status of the target. Looking at the simple main effects of region of interest at different levels of status, the trend for the eyes to be most frequently fixated followed by the face and then the body was the same in both high- ($F(2,23)=52.2$, $p<.001$) and medium-status targets ($F(2,23)=77.3$, $p<.001$). In each case, comparisons between the different regions of interest were all reliable (at least $p<.05$). There was also an effect of region of interest in the low social status target, $F(2,48)=31.7$, $p<.001$. In these targets, there was still a tendency to fixate the eyes rather than the face or body (both $p<.001$). However, unlike in the other targets, there was no reliable difference between the likelihood of looking at the face compared to the body.

Discussion

This experiment explored the spatiotemporal distribution of gaze in a controlled but realistic video of a social interaction. Unlike the vast majority of research into social attention, we used stimuli containing several individuals conversing in a dynamic situation (a video), and this allows us to draw some conclusions about how visual attention is directed in complex scenes with a truly social element. The evolutionary research reviewed in the introduction leads to the straightforward predictions that humans should be predisposed to attend to other people in the environment, to their eyes (Emery, 2000), and to high-status people in particular (Henrich & Gil-White, 2001). Testing these predictions led to several interesting findings.

First, people chose to spend a majority of the time looking at the people in the clips, even though these people did not occupy the entire scene. Of course this is not all that surprising considering that the other regions in the movie (background and furniture) were motionless, not useful for the task, and probably not as salient in terms of low level features, but it does confirm previous reports that the visual attention system is particularly inclined to select people, and extends these findings to video. More interesting, most of the fixations on people were targeted at an individual's eye region, with fewer gazes directed at the rest of the face, and fewer still at the torso and other body parts. Participants spontaneously chose to monitor the eyes of the people in the clips, and this extends to natural dynamic scenes what has previously only been found for static images (Birmingham, et al., 2008) and Hollywood movies (Klin, et al.,

2002). It is not known why the eyes are so potent at drawing attention. The fact that humans have uniquely salient eyes (due to higher contrast between the dark iris and white sclera than in any other primates) suggests that we are physiologically evolved to communicate our eye gaze direction to others (see Emery, 2000, for this and other examples of how humans are physiologically specialized for this task). On the other hand, the tendency for people to look at eyes does not appear to be due solely to their low-level conspicuity (Birmingham, Bischof, & Kingstone, 2009). Instead, our results add to a body of data suggesting that people are predisposed to look at the eyes because they have evolved to be particularly informative about the beliefs, intentions and goals of other agents. In our naturalistic task, watching a social interaction while thinking about some of the people involved, the eyes were spontaneously selected as especially useful.

Second, a range of different measures demonstrated that the relative social status of the people in the clips had a large and robust effect on who was fixated. People who were previously rated as having high social status—whom other group members perceived as having led the task or influenced the group—were fixated more often, for longer on each gaze, and for a longer total time, compared to people seen as medium social status, or low social status, and low-status targets received the least attention. The independently rated status hierarchy of the group depicted in the videos had a highly systematic effect on the distribution of gaze of participants watching the clips. Why did social status affect how much a person was looked at? Although both the position of a person in the scene and their verbalizations had an effect on the

amount of attention they received, our analyses indicate that the effect of social status could not be attributed to either of these factors. This was clear in multiple different analyses. For example, high-status people were looked at more often than medium-status people whether they were positioned in the centre of the group or on the sides. High-status targets spoke slightly more often than medium-status targets (although this difference was not statistically significant), but the effect of the social status hierarchy on attention held even in those moments when nobody (or somebody else) was talking, and when variance in speaking time was statistically removed. Our eye-tracking methodology allowed us to look in detail at the gazes that each target received, and the three measures reported can reveal slightly different aspects about the bias shown towards high-status targets. The fact that participants spent a greater amount of total time looking at these targets—often several seconds more within a short 20s clip—could be attributed to a higher frequency of shifts toward these people or a longer time spent looking at them each time they were there. In fact, both these patterns were found, with participants making *more* fixations on high-status targets as well as *longer* gazes. These findings demonstrate that people are more likely to shift their gaze to high-status targets, and that once there they stay there for longer before looking at someone else.

The strong effects of social status are particularly interesting given that participants saw only brief episodes of the social interaction in each group. One explanation of the high-status advantage is that status was inferred from aspects of the targets' appearance, from their non-verbal behaviour, or from other group members'

behaviors and responses toward them (e.g., others asked them for advice, deferred to their opinions, etc.). High-status targets may have been physically larger, sat more expansively or acted more dominantly in some way (Aries, Gold, & Weigel, 1983; Cashdan, 1998). Indeed, in the group task described here several of these behaviours were found to distinguish between high- and low-status people (Cheng, Tracy & Henrich, 2010; Cheng et al., in prep), an observation predicted by evolutionary theory (Henrich & Gil-White, 2001). It is likely that such cues, along with the increased tendency to speak and the content of what was said are what lead some individuals to be rated as high-status by their peers and other observers. Further research is necessary to identify whether these behaviours are in themselves salient attractors of attention, or whether an attribution of status is necessary. In the case of speaking, our findings demonstrate that high-status targets were looked at more often than medium- and low-status targets even when somebody else, or nobody, was talking, suggesting that it is their status within the hierarchy, rather than their verbal behaviour at that time, that results in them being paid the most attention. The implication is that people can very quickly ascertain who the high-status members of a group are, and that they are predisposed to orient toward these people. This attentional bias may represent an evolved cognitive mechanism that facilitates the detecting and monitoring of high-status individuals (Henrich & Gil-White, 2001). Increased attention toward these individuals might allow group members to appropriately monitor the goals and behaviors of their leaders, learn from these individuals, who tend to possess superior skills, and monitor potential threats or attacks from these more powerful conspecifics. These data also lend

strong support to the idea that others could use gaze following as an indicator of social status within a group: the person who receives the most glances from other group members, or who is gazed at the longest may be perceived as the high status individual (Chance, 1967; Emery, 2000).

By looking at the temporal synchrony between who was speaking and who was being looked at, our experiment also addressed the relationship between gaze and speech. A significant body of research has investigated where people fixate when observing someone talking. Somewhat surprisingly, in general both humans (Vatikiotis-Bateson, Eigsti, Yano, & Munhall, 1998) and monkeys (Ghazanfar, Nielsen, & Logothetis, 2006) look mostly at the eye region, rather than the mouth region, of a vocalizing conspecific. This pattern is confirmed in our finding that the eyes were indeed looked at most frequently. When auditory noise is added to the speech, or when the task requires accurate auditory discriminations, a higher frequency of fixations are made to the mouth (Buchan, Pare, & Munhall, 2007). However, the vast majority of these studies displayed the face of a single speaker performing a monologue, rather than the real conversation used here.

In research involving more interactive communication, Richardson, Dale and Kirkham (2007) have documented the "gaze-coordination" in a conversation: conversants tend to look at the same thing at the same time. Other descriptions of the role of gaze in conversation suggest that it functions as a social signal for whose turn it is to talk next (Kendon, 1967). In our own analysis, we found that observers were quite likely to look at the person talking at any one moment (and most of the time this was at

their eyes), but that gaze tended to predict the change from one speaker to the next. A similar finding was recently reported by van Hofsten et al (2009), who analysed the proportion of saccades that went from one speaker to the other within 2 seconds of the change in speaker. This study found that normally functioning children made these turn-tasking gaze shifts frequently, but that they were significantly less common in children with autistic spectrum disorder. In our study, we used a novel technique (cross-correlation) to quantify the temporal lag between gaze and speech, and we found that the observing participants tended to look at targets around 150ms before they spoke. Obviously our participants did not have the opportunity to actually converse with the targets, but it may be that the temporal pattern in gaze shifts reflects the general pattern of turn-taking during a conversation. That participants' gaze predicted who was going to speak next may indicate that the next speaker was being addressed or referred to by the current speaker (and so the observer may have been looking to observe their reaction), or that the context constrained who was going to speak next in other ways. This intriguing finding merits further study.

In conclusion, we have used a complex, realistic and social stimulus to explore the allocation of gaze in a group interaction. The people in this interaction, and in particular their eye regions, were potent targets for fixation. However, high-status individuals were looked at more often and for longer than low-status targets, which is consistent with a rapid perception of the social hierarchy in the scene and an evolutionarily determined bias toward attending to some people more than others. Gaze was also temporally yoked to the conversation between the people. These

findings are among the first to demonstrate the influence of a realistic social context and the hierarchy that goes with it on the top-down allocation of eye gaze, and they provide a way forward for researchers investigating social attention.

References

Abramovitch, R. (1976). The relation of attention and proximity to rank in preschool children. In M. R. A. Chance & R. R. Larsen (Eds.), *The social structure of attention*. (pp. 153-176). London: Wiley.

Aries, E. J., Gold, C. & Weigel, R. H. (1983). Dispositional and situational influences on dominance behavior in small groups. *Journal of Personality and Social Psychology*, 44 (4), 779-786.

Ballard, D., & Sprague, N. (2005). Modeling the brain's operating system. *Brain, Vision, And Artificial Intelligence, Proceedings* (Vol. 3704, pp. 347-366).

Baron-Cohen, s. (1995). *Mindblindness: an essay on autism and theory of mind*. Cambridge(MA): MIT Press.

Berger, J., Rosenholtz, S. J., & Zelditch, M. (1980) Status organizing processes. *Annual Review of Sociology, 6*, 479-508.

Boehm, C. (1993). Egalitarian society and reverse dominance hierarchy. *Current Anthropology, 34*, 227-254.

Birmingham, E., Bischof, W. F., & Kingstone, A. (2008). Gaze selection in complex social scenes. *Visual Cognition, 16*(2-3), 341-355.

Birmingham, E., Bischof, W. F., & Kingstone, A. (2009). Get real! Resolving the debate about equivalent social stimuli. *Visual Cognition, 17*(6-7), 904-924.

Buchan, J. N., Pare, M., & Munhall, K. G. (2007). Spatial statistics of gaze fixations during dynamic face processing. *Social Neuroscience, 2*(1), 1-13.

Cashdan, E. (1998). Smiles, speech and body posture: How women and men display sociometric status and power. Journal of Nonverbal Behavior, 22(4), 209-228.

Chance, M. R. A. (1967). Attention Structure as Basis of Primate Rank Orders. *Man, 2*(4), 503-518.

Cheng, J. T., Tracy, J. L., & Henrich, J. (2010, January). Are dominance and prestige distinct strategies for attaining social status? Poster presented at the annual meeting of the Society for Personality and Social Psychology. Las Vegas, Nevada.

Dalton, K. M., Nacewicz, B. M., Johnstone, T., Schaefer, H. S., Gernsbacher, M. A., Goldsmith, H. H., et al. (2005). Gaze fixation and the neural circuitry of face processing in autism. *Nature Neuroscience, 8*(4), 519-526.

Emery, N. J. (2000). The eyes have it: the neuroethology, function and evolution of social gaze. *Neuroscience And Biobehavioral Reviews, 24*(6), 581-604.

Findlay, J. M., & Gilchrist, I. D. (2003). *Active Vision: The Psychology of Looking and Seeing* Oxford: OUP.

Foulsham, T., & Underwood, G. (2007). How does the purpose of inspection influence the potency of visual saliency in scene perception? *Perception, 36*, 1123-1138.

Foulsham, T., & Underwood, G. (2008). What can saliency models predict about eye movements? Spatial and sequential aspects of fixations during encoding and recognition. *Journal Of Vision, 8*(6), 1-17.

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

Ghazanfar, A. A., Nielsen, K., & Logothetis, N. K. (2006). Eye movements of monkey observers viewing vocalizing conspecifics. *Cognition, 101*(3), 515-529.

Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., & Malach, R. (2004). Intersubject synchronization of cortical activity during natural vision. *Science, 303*(5664), 1634-1640.

Hayhoe, M., & Ballard, D. (2005). Eye movements in natural behavior. *Trends In Cognitive Sciences, 9*(4), 188-194.

Henderson, J. M. (2003). Human gaze control during real-world scene perception. *Trends In Cognitive Sciences, 7*(11), 498-504.

Henrich, J. & Gil-White, F. (2001) The Evolution of Prestige: freely conferred status as a mechanism for enhancing the benefits of cultural transmission. *Evolution and Human Behavior*, 22, 1-32

Itti, L. (2005). Quantifying the contribution of low-level saliency to human eye movements in dynamic scenes. *Visual Cognition, 12*(6), 1093-1123.

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research, 40*(10-12), 1489-1506.

Kendon, A. (1967). Some Functions of Gaze-Direction in Social Interaction. *Acta Psychologica, 26*(1), 22-&.

Kingstone, A. (2009). Taking a real look at social attention. *Current Opinion In Neurobiology, 19*(1), 52-56.

Klin, A., Jones, W., Schultz, R., Volkmar, F., & Cohen, D. (2002). Visual fixation patterns during viewing of naturalistic social situations as predictors of social competence in individuals with autism. *Archives of General Psychiatry, 59*(9), 809-816.

LaFreniere, P. J., & Charlesworth, W. R. (1983). Dominance, attention, and affiliation in a preschool group: A nine-month longitudinal study. *Ethology and Sociobiology 4,* 55-67.

Land, M. F., & Hayhoe, M. (2001). In what ways do eye movements contribute to everyday activities? *Vision Research, 41*, 3559-3566.

Maner, J. K., DeWall, C. N., & Gailliot, M. T. (2008). Selective attention to signs of success: Social dominance and early stage interpersonal perception. *Personality and Social Psychology Bulletin, 34*(4), 488-501.

McNelis, N. L., & Boatright-Horowitz, S. L. (1998). Social monitoring in a primate group: the relationship between visual attention and hierarchical ranks. *Animal Cognition, 1*, 65-69.

Pylyshyn, Z. W., & Storm, R. W. (1988). Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial Vision, 3*, 179 - 197.

Raymond, J. E., Shapiro, K. L., & Arnell, K. M. (1992). Temporary Suppression of Visual Processing in an Rsvp Task - an Attentional Blink. *Journal Of Experimental Psychology-Human Perception And Performance, 18*(3), 849-860.

Richardson, D. C., Dale, R., & Kirkham, N. Z. (2007). The art of conversation is coordination - Common ground and the coupling of eye movements during dialogue. *Psychological Science, 18*(5), 407-413.

Rogers, R. D., & Monsell, S. (1995). Costs of a Predictable Switch between Simple Cognitive Tasks. *Journal Of Experimental Psychology-General, 124*(2), 207-231.

Shepherd, S. V., Deaner, R. O., & Platt, M. L. (2006). Social status gates social attention in monkeys. *Current Biology, 16*(4), R119-R120.

Shepherd, S. V., & Platt, M. L. (2008). Spontaneous social orienting and gaze following in ringtailed lemurs (Lemur catta). *Animal Cognition, 11*(1), 13-20.

Vatikiotis-Bateson, E., Eigsti, I. M., Yano, S., & Munhall, K. G. (1998). Eye movement of perceivers during audiovisual speech perception. *Perception & Psychophysics, 60*(6), 926-940.

von Hofsten, C., Uhlig, H., Adell, M., & Kochukhova, O. (2009). How children with autism look at events. *Research in Autism Spectrum Disorders, 3*(2), 556-569.

Vaughn, B. E., & Waters, E. (1981). Attention structure, sociometric status, and dominance: Interrelations, behavioral correlates, and relationships to social competence. *Developmental Psychology, 17,* 275-288.

Yarbus, A. L. (1967). *Eye movements and vision*. New York: Plenum.