Corresponding Author: Dr. Tom Foulsham,

Corresponding Author's Institution: University of British Columbia

First Author: Tom Foulsham

Order of Authors: Tom Foulsham; Joey T Cheng; Jessica L Tracy; Joseph Henrich; Alan Kingstone

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.

Department of Psychology,
University of British Columbia,
2136 West Mall,
Vancouver, B.C.
V6T 1Z4
Canada


8 September 2010


Ms. Ref: COGNIT-D-10-00084


Dear Dr. Hollingworth;

Thank you for the reviews and for the opportunity to revise and resubmit our manuscript for *Cognition,* now entitled "Gaze allocation in a dynamic situation: Effects of social status and speaking".

As you may recall, both reviewers liked the original submission and recommended its publication after relatively minor revisions. Specifically, the reviewers made several helpful comments that were addressable without collecting further data. We have addressed all of their comments in the new version, and we include a point-by-point response below.

In addition, you made an interesting suggestion that our participants might have looked more at the high-status targets because they were oriented toward the goal we had given them in the instructions, of evaluating the targets' effectiveness. To address this possibility we followed your advice to perform a control study, with a manipulation of the instructions.   Specifically, as you suggested, we asked participants to think about who they would *not* want to work with (i.e. who they would avoid in a subsequent task).  Despite this change of instructions, results revealed the same bias for gazes toward high-status individuals (reported on p. 19 of the revised manuscript). This replication suggests that the effects of social status on gaze are not merely caused by people looking towards the target implied by the instructions. We would like to thank you for raising this possibility in your action letter, as these new data have clearly helped strengthen the conclusions that can be drawn from our study.

As requested, we have marked our changes in red font. We believe the revised manuscript is substantially improved, and we look forward to hearing from you.


Yours sincerely,


Dr. Tom Foulsham

Ms. Ref. No.:  COGNIT-D-10-00084

**Reviewers' comments (<span style="color:red">authors' response in red</span>)**


Reviewer #1:

Summary and Evaluation:  This paper presents detailed analyses of the fixation patterns of participants who viewed a series of three-person interactions in which the 3 people differed in social status. Analyses of the videos indicated that persons of higher social status tended to talk more than the other people in the video. Participants who viewed the videos tended to look more at the persons of higher social status (on several measures), even when speaking time and spatial location were accounted for. Interestingly, in cross-recurrence analyses of participant gaze and the speech of individuals in the videos (e.g., Richardson & Dale, 2005), participant viewers tended to shift their gaze towards the next speaker about 150ms prior to speech onset. This finding suggests that in these situations, participants were able to anticipate who would speak next.

    I found this paper to be very interesting. In particular, the result that I find most interesting is that participants looked at the speakers before they started talking. As I mention in my small comments below, other work in the language comprehension literature finds that listeners tend look at what's being talked about *after* it has been mentioned, with some notable exceptions in natural conversation, and in situations where the language allows the listener to predict upcoming words. The present result is different for a variety of reasons-the participants were simply watching a video of other people talking abstractly, and also the relevant gaze was directed at the speaker, not what was mentioned (as they were talking about a hypothetical situation, not objects on a tabletop). The finding that social status predicts gaze at the people in the video is really neat. The authors draw links with evolutionary theory to explain this finding, as well as the finding that participants tended to look at the eyes, both of which are interesting ideas. My biggest suggestions are: (1) Better explain how the -150ms cross-correlation effect fits in with other findings in the gaze and language comprehension literature, which show that fixations typically lag behind speech

<span style="color:red">We have added to the discussion to address this point (see reviewer's details and our reply below).</span>

(2) clarify a variety of small points about the analyses which were done. Overall, I found this to be a really interesting paper, and I think that it should be published with only minor revisions.

Small points

p.11:  In the discussion of how the video clips were selected, it is stated that clips in which the participants were negotiating and reaching a pivotal decision were selected because they captured places in which the status dynamics were most salient. I wonder if the authors could say a bit more about why the status dynamics were most apparent here. Also, if they could provide some insights (even intuitions) about what made one person have a higher status than the others, would be great. The general discussion touches on this briefly (p.32) but a somewhat extended discussion would be interesting.

We have added material on p.11 describing why we intuit there are status differences here.  Specifically, we now say:

> "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.  These clips often contained moments of conflict, or times when one target had succeeded in persuading other members after an extensive debate, thus they captured moments where status dynamics were particularly salient. "

Further details about the statistical differences between the targets are also given on pp.10-11 (see Reviewer 2's comment below).

In addition, on p.35, we have extended our discussion of the determinants of social status, and the implications of our findings for our understanding of status dynamics.

p.11-12:  From a purely methodological point of view, the way in which you solved video-tracker timing issue is likely going to be of interest to a lot of readers who do eye-tracking work. It'd be great if you could say more about how you got the Eyelink data file to synch with the audio/video code-Did you use the psychophysics toolbox in Matlab or some other software? Did you pass information about when the audio and video stim began to the Eyelink data file to synch them or was the synching done by some other means?

We have described the method of syncing in greater detail on p.12:

"The dynamic nature of our stimuli meant that maintaining a temporal synchrony between video, audio and eye tracking data was important. We accomplished this using EyeLink's standalone Experiment Builder software, which, in concert with the Xvid codec, wrote with millisecond accuracy time-stamped messages to the eye tracking data file at the onset (i.e. the first screen retrace) of each video frame. Several frames were cached in advance and the system demonstrated extremely accurate timing. The timestamp for each frame could then be compared to the events in the eye tracking data file (e.g. fixations)."

p.13: When "targets" is mentioned here I was a little confused if this meant the humans (which I think it does) or included other objects in the scene. If a more descriptive name could be used, or a little more description just at the beginning of the section, that would be great.

We use the word target to refer to the humans in the video clips, to maintain consistency with research in social and personality psychology on social status and human behavior more generally. We have added a more explicit definition of the term "target" both in the method section, where it is first mentioned (p.9), and at the beginning of the results (p. 14).

p.15: Regarding the ANOVA on the proportion of fixations: To be clear, I do not think the authors need to change their analyses (I'll explain why after an aside). But, just as an FYI, there is increasing discontent regarding the use of ANOVA to analyze proportional data, because proportions are bounded at 0,1 (as well as a variety of other reasons). A special issue of JML in 2008 has a number of papers with treatments of this issue; I suggest taking a look at Jaeger (2008). Some folks have suggested in the past using an arcsine transformation of proportions prior to ANOVA, however Jaeger points out some flaws with this solution. Unfortunately, for the analysis of eye-tracking data there isn't a great solution yet, though a number of folks have suggested potential solutions (see: Barr, 2008, Mirman, Dixon, & Magnuson, 2008; McMurray, Samelson, Lee, & Tomblin, 2009; Dunabeitia, Aviles, Afonso, Scheepers, & Carreiras, 2009). Most of those papers are dealing with how to model
changes in fixation proportion (or odds) over time, which is largely unrelated to the present paper. Unfortunately, each of these proposed approaches to analyzing eye-tracking data has major drawbacks, and none has really caught on yet in the (visual world/language processing eye-tracking) field, nor have they been thoroughly vetted. Thus, as I said at the beginning of this aside, I don't think for now the authors

should do anything differently, largely because it is not clear that anything much better can be done, I just wanted to point out this issue.

We are grateful for these statistical references and agree that other methods are worth developing for eye tracking data. One of the advantages of our analysis is that we also look at total fixation time (which is not bounded by 0,1 and therefore more amenable to analysis of continuous variables).

One other question for the authors on this point: For the proportion of fixation calculations, was a single fixation defined as a single data frame in which there was a fixation on some object, as a series of frames in which there was a fixation on the same object, or some other way?  i.e., Were the proportions calculated based on the ~5000 dataframes for each clip, or the 3-8 fixations made during each clip? If it were the latter, one might expect there to be a SMALLER proportion of fixations on the high-status targets, as each individual fixation lasted much longer.

We have added a clarification about this on p.15.  Fixations were defined, as is standard in eye tracking research, according to velocity and acceleration thresholds. As we say at the start of the results, participants made 49 fixations on average during each 20s (600 video frames) clip.  Thus, the proportion of fixations gives a measure of the total information processing spent on each target, allowing for any differences in the total number of fixations.  Often multiple consecutive fixations were made on a target (e.g. on a different part of the face) before moving on (what we and others call a "gaze").

p.19:  When it's stated that the difference in talking time between high and medium status targets wasn't significant, I was left wondering whether it might have been a marginal effect, or significant if a one-tailed test had been used. Thus it'd be great if the t-value could be reported here.

We now report the t and p values at this point (now p.21), and at p=.19 it is not a marginal effect.

p.20:  It would be great if the effect sizes for the effects of position, time, and social status could be reported. Alternatively, if the authors were to take a mixed-model logistic regression approach (e.g., Jaeger, 2008), the relevant data would be the estimated beta weights for the three factors.

We have added effect sizes (partial eta-squared values) for all the ANOVA/ANCOVA effects in the paper.

p.22: What is meant by eye-guidance?
     Also, I'm a little confused by the dependent measure used in the analyses at the bottom of the page. The first paragraph seems to indicate that these analyses only focus on the first frame of a fixation on a target. But then in the second paragraph, it says that "high-status person was talking, target status had a reliable effect". This seems to imply that as they were talking, most of the fixations were on them, thus the dependent measure included all fixations. A little clarification would be helpful.

The analyses at this point include all fixations, as per the definition given earlier. Our point in adding the clarification about the first frame is that, because targets may have stopped/started talking *during* a fixation, whether that fixation is on a talking person is ambiguous. However, at this point we want to know whether, when the saccade that brings the eyes to the target is planned and initiated, the target of that saccade was talking (for example, target 1 starts talking, the observer detects the speaker and moves the eyes from elsewhere to that target).  This is what we mean by eye *guidance*.  We have added clarification of this point in this revision (p.24):

> "At this point we were interested in how the destination for each fixation was planned or guided, so we categorized 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. We analysed all fixations, and made a comparison between the proportion of these fixations that were directed at the person talking (at the start of that fixation) and those that were on another target or on the scene background."

p.26: It was really surprising to me that the cross-recurrence analysis showed the best correlation at -5 frames, as Richardson & Dale (2005, also Richardson, Dale, & Kirkham, 2007) have observed an approximately 2s delay in the coordination of speaker and listener's eye movements. And, in carefully controlled studies of speech perception using gaze (e.g., Allopenna, et al., 1998), it is found that it takes approximately 200-300ms following a critical word onset to fixate an associated picture in the scene. I wonder if the divergence from these findings has to do with the fact that the participants in the present study were fixating faces rather than mentioned objects, or watching a dialog from an overhearer perspective (see Schober & Clark, 1989). Or, if it

was the fact that the participants in the present study didn't have a task to follow. The observed result, then, suggests that participants were able to anticipate who would talk next, which is quite interesting. I think the observed results may fit nicely with findings by Brown-Schmidt & Tanenhaus (2008) who found that addressees in a conversation very often were fixating a to-be-mentioned target well before the speaker ever mentioned it, possibly because pragmatic information related to the task predicted what would be talked about next. It also fits with findings that when the language itself is predictive of what will be mentioned next, listeners make anticipatory fixations to a to-be-mentioned object (Altmann & Kamide, 1999). It would be worthwhile to try to compare and contrast the observed findings with the previous literature mentioned, given how (possibly) surprising it is.

We have expanded the discussion in order to address this result much more thoroughly (starting p.39).  As you will see, the reviewer's references to previous studies were very helpful.  The difference between past studies and our own results seems to be that in our task, unlike most others (e.g. in the visual world paradigm), people were not talking about concrete objects which could be gazed at. In fact, the target of gaze (i.e. the people in the clip) were (typically) not being explicitly cued by the contents of the dialog. We suggest that in our study people are following the dynamic turn taking of the conversation and using patterns in speech and gesture to predict who will speak in the next few frames. As the reviewer says, this is consistent with other evidence that people can anticipate mentioned items given pragmatic and linguistic context.

        Finally, it is mentioned on this page that the -150ms offset was found for all 3 types of target speaker, but it would be nice if some statistics could be presented. I imagine the analyses might be hard to do as there are more observation opportunities for high and mid-status talkers (as they talked more), but due to this, it might be the case that participants are better at anticipating when high-status talkers will begin speaking than low-status talkers (again here is a case where a mixed-model logistic regression approach would be helpful).

We now report the median lag separately for each target type (now p.28), although the implications of these data are suggestive at best, for the reason noted by the reviewer and because we only had a few examples of each target type.

p.33: Regarding the phenomenon of looking at the eyes of a speaker; I believe this is the case as well in sign languages.

<span style="color:red">This is an intriguing result, and we now cite an empirical study reporting it on p.38. We are grateful to the reviewer for pointing it out.</span>

p.34:  Given that Richardson and Dale have used cross-correlation analyses for quite some time now, I don't know that I would call this application of cross correlation analyses a "novel technique", though the result is, I believe, novel.

<span style="color:red">We have reworded this sentence, and we also now mention Richardson and Dale's paper which used the technique first.</span>

Reviewer #2: Review of "Gaze allocation in a dynamic social situation: Effects of social status and speaking" (COGNIT-D-10-00084)

The authors conducted an eyetracking experiment to test the hypothesis that people attend preferentially to people who appear high in status more than people medium and low in status. I found much to like about this paper. The method was elegant, the results were thoroughly supportive of the authors' hypotheses, and the conceptual framework was sufficiently strong to make a novel contribution to several literatures (e.g., power, evolutionary psychology, eye movements). I do have some concerns that I think could be handled in a revision without the collection of additional data.

My first concern is that the authors did not do a sufficient job explaining why people, including the people in their sample, attend preferentially to targets' eyes and faces. The authors only mention that people tend to do this, but they don't provide a compelling theoretical explanation why people do this. As a result, the reader is left without much to "chew on" in terms of the mechanism underlying people's preferential attention to eyes and faces. Mark Dadds has done some recent eyetracking work showing deficits in processing of eyes among children who show signs of psychopathy. I would check out his and others' papers to learn more about the function of attentional processing to eyes. I strongly encourage to strengthen this component of their theoretical model.

We have added material on the eye bias; specifically discussing the possible function of this bias and neuropsychological findings, such as that mentioned by the reviewer.  Our revised discussion of this begins on p.32:
"Humans have physiologically evolved to communicate their eye gaze direction to others (for example by having a high contrast between the iris and sclera see Kobayashi & Koshima, 1997), but the function of attentional orienting to the eyes is not fully understood. Evidence suggests, however, that the ability to share others' goals and intentions (i.e., theory of mind) is fundamentally linked to orienting to another's gaze (Tomasello, Carpenter, Call, Behne, & Moll, 2005).  For example, evidence from individuals with autism (Baron-Cohen, 1995), psychopathy (Dadds, El Masry, Wimalaweera, & Guastella, 2008), and patients with selective damage to the amygdala (Adolphs, et al., 2005), confirms that deficits in processing emotion and theory of mind are often accompanied by a reduced tendency to look at the eyes.  The hormone oxytocin enhances both social approach behaviour and fixations to the eyes in humans, confirming this link and suggesting part of the mechanism (Guastella, Mitchell, & Dadds, 2008).  In our naturalistic

task, watching a social interaction while thinking about some of the people involved, the eyes were spontaneously selected by observers as being especially useful.  This is consistent with a theoretical model that attentional processing of the eyes functions to enhance the perception of the target's external goals (e.g. who they are talking to) as well as their internal emotions, intentions and beliefs (e.g. how they feel about that person; see Birmingham & Kingstone, 2009 for a recent review)."

My second concern is that the authors need to provide more information about the pre-testing of their video clips. At minimum, the authors should report the results of statistical analyses to show that the high-status targets were significantly higher in perceived status compared to medium-status targets, who were in turn significantly higher in perceived status compared to low-status targets.

We have added material on pp. 10-11 explaining how we determined that the targets included differed in status from each other, and could be used to represent each assumed status level. Specifically, we now say:

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.  These ratings were significantly different ($F(2, 9)=7.7$, $p<.05$).  Planned contrasts across the in-group ratings of the four examplars of each target (i.e., across the four 'high-status' targets, the four 'low-status' targets, and the four 'medium-status' targets) confirmed that low status targets were rated significantly lower in status than high ($t(9) = 3.83$, $p<.01$) and medium ($t(9)=2.63$, $p<.05$) status targets.  High and medium targets did not differ significantly.  As a further manipulation check, therefore, we asked our eye-tracked participants (after the experiment), and an additional 34 naïve raters, to rate the social status of each of the targets (using the same three items and 7-point scale as the in-group ratings).  These participants rated our three groups of targets as significantly different ($F(2, 174)=110.46$, $p<.001$, $\eta_p^2 = .56$).  High status targets (M=4.97, SD=1.20) were rated higher than medium status targets (M=4.21, SD=1.06), who in turn were perceived as having significantly higher status than low status targets (M=2.25, SD=0.76).  Contrasts between these levels were all highly reliable (all $ts(174)>4$, $ps<.001$).

My third concern is that the authors did not measure any individual difference variables that may have moderated the observed effects. For example, would attention to high-status targets be strongest among participants who generally occupy positions of power and leadership? Would they be strongest among participants high in social dominance orientation? Are the effects moderated by participant gender? I encourage the

authors to bring this up in the Discussion as a potential avenue for future research.

We followed the reviewer's advice here, and have added predictions and potential future research directions regarding individual differences that may moderate these effects, in the discussion (p.36, last paragraph).

My fourth concern is that the authors did not assess how targets' attractiveness influenced attentional bias. This will strengthen the paper in at least two ways. First, attentional bias to attractive static images is a well-replicated finding (e.g., Maner et al., 2003, 2007, 2008), but it hasn't been studied much with dynamic images. Second, the authors would be able to determine whether their effects are distinct from targets' attractiveness. I encourage the authors to code the attractiveness of targets and re-analyze their data appropriate to examine the effect of attractiveness on their effects.

We agree that attractiveness is likely to be an important factor, and one that merits study. However, in considering the reviewer's suggestion, we realized that in the present case this is a complex issue. Given our use of dynamic stimuli, for example, it is unclear whether the most appropriate measure would be raters who see static images of the targets (which might indicate their "physical" attractiveness) or raters who see the video clips (which would include gestural and dialog cues so might be something quite different)? While the latter would be more similar to what participants did, it would also allow for spontaneous status judgments to be made, which could influence attractiveness ratings. Thus, while we agree with the reviewer that this idea should be taken seriously, we feel that with the current study design there is no clear-cut way to address it, and any attempt to do so could promote conclusions that may not be warranted.
        Nonetheless, we were able to use one piece of evidence, from the existing data set, which suggests that attractiveness is far from the entire story. Amongst the battery of judgements made about the targets, we included the item, "I like the target".  Scores on this item were a predictor of attention, but this was a small effect, and it was reduced to non-significance when speaking was controlled for. Furthermore, controlling for this item did not affect the robust relation between targets' status and attention. We have added material discussing these considerations on p.37.

In sum, I enjoyed the paper quite a bit. I hope that these comments assist the authors as they pursue this line of work.

# Gaze allocation in a dynamic 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; speech perception; dynamic scene perception

Word count: 9614

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, Jones, Schultz, Volkmar, & Cohen, 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, et al., 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 20 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
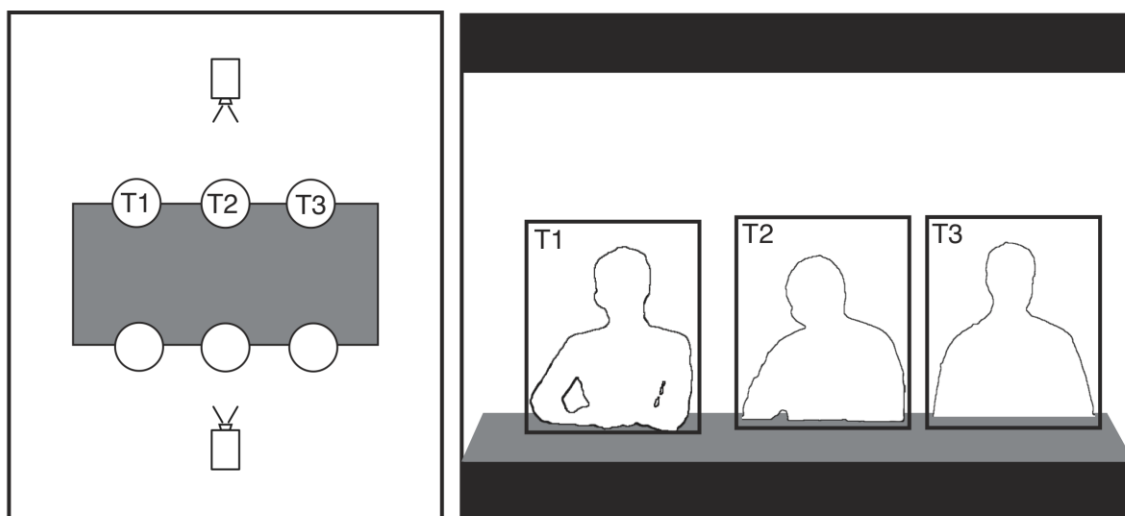
the layout of the resulting video frames.



**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.

Four representative videos were chosen for the eye tracking study. In each case, the

three individuals in the video were be classified according to social status scores

from the original group-interaction experiment (from now on we will refer to these

three individuals as the "targets"). 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. These ratings were significantly different ($F(2, 9)=7.7$, $p<.05$). Planned contrasts across the in-group ratings of the four examplars of each target (i.e., across the four 'high-status' targets, the four 'low-status' targets, and the four 'medium-status' targets) confirmed that low status targets were rated significantly lower in status than high ($t(9) = 3.83$, $p<.01$) and medium ($t(9)=2.63$, $p<.05$) status targets. High and medium targets did not differ significantly. As a further manipulation check, therefore, we asked our eye-tracked participants (after the experiment), and an additional 34 naïve raters, to rate the social status of each of the targets (using the same three items and 7-point scale as the in-group ratings). These participants rated our three groups of targets as significantly different ($F(2, 174)=110.46$, $p<.001$, $\eta_p^2 = .56$). High status targets

(M=4.97, SD=1.20) were rated higher than medium status targets (M=4.21, SD=1.06), who in turn were perceived as having significantly higher status than low status targets (M=2.25, SD=0.76).  Contrasts between these levels were all highly reliable (all ts(174)>4, ps<.001).

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 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.  These clips often contained moments of conflict, or times when one target had succeeded in persuading other members after an extensive debate, thus they captured moments where status dynamics were particularly 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.

The dynamic nature of our stimuli meant that maintaining a temporal synchrony between video, audio and eye tracking data was important.  We accomplished this using EyeLink's standalone Experiment Builder software, which, in concert with the Xvid codec, wrote with millisecond accuracy time-stamped messages to the eye tracking data file at the onset (i.e. the first screen retrace) of each video frame.  Several frames were cached in advance and the system

demonstrated extremely accurate timing.  The timestamp for each frame could then be compared to the events in the eye tracking data file (e.g. fixations).

**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 three people in the clip (i.e., the "targets") 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. Fixations were parsed by the EyeLink system, according to the velocity and acceleration thresholds outlined in the method, and proportions were calculated across all fixations made during a clip, and then averaged across clips.

| | 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 ($F(2,48)=31.7$, $p<.001$, $\eta_p^2=.57$). 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 ($F(2,48)=34.1$, $p<.001$, $\eta_p^2=.59$). 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, $\eta_p^2$=.35). 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

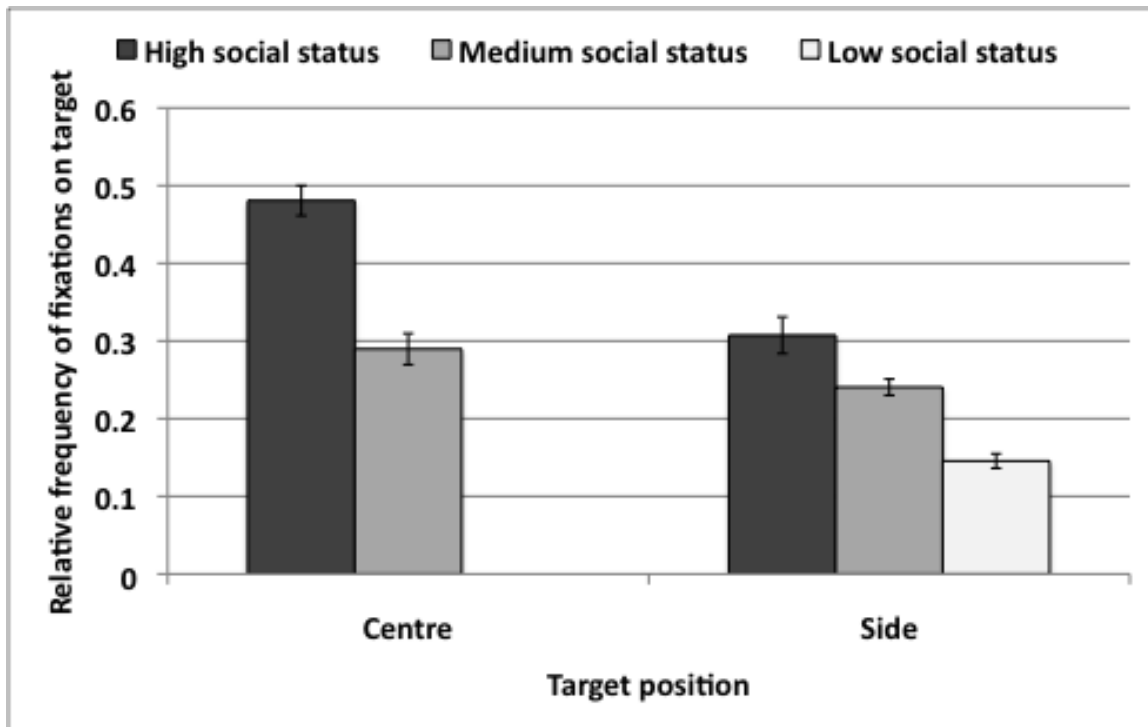display), social status had an impact on attention at all seating positions.

**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.

One question raised by these results is whether the effect of social status on eye gaze is spontaneous, or partially due to our instructions to eye-tracked participants, to "think about who you would want to work with on a subsequent task"; it is possible that these instructions encouraged observers to look at the high-status individuals. Although this issue does not change our main finding, that individuals *can* automatically orient their gaze toward high status individuals and determine which individuals in a group interaction are high-status, despite viewing these individuals for only very brief time periods, we conducted a control experiment to determine whether this process is spontaneous or potentially goal-directed. We tested an additional 7 participants (with the same characteristics as

those in the main study) as they watched one of the four sets of clips, and we instructed these observers to "think about who you would NOT want to work with (i.e., who you would want to avoid working with) on a subsequent task".  All other parts of the experiment were unchanged.  The results showed that there was still an effect of social status on the proportion of fixations ($F(2,12)=4.8$, $p<.05$, $\eta_p^2=.45$) and the total fixation time ($F(2,12)=6.2$, $p<.05$, $\eta_p^2=.51$), and that in each case the high social status target was prioritized over the medium target, who was in turn gazed at more than the low target (all $ps<.05$).  Thus, even when instructed to think about the least effective group member, observers spontaneously selected targets according to social status. These findings indicate that prioritizing high status individuals with eye gaze is a spontaneous rather than induced behavior.

**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, $F(2,71)=11.4$, $p<.001$, $\eta_p^2=.25$). 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 ($t(46)=1.3$, two-tailed $p=0.19$). 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 ($F(1,67)=12.3$, $p<.005$, $\eta_p^2=.15$) and talking time ($F(1,67)=35.9$,

$p<.001$, $\eta_p^2=.35$) 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 ($F(2,67)=16.8$, $p<.001$, $\eta_p^2=.33$). The same hierarchy of attention was

seen, with high-status targets being fixated more often than medium-status targets

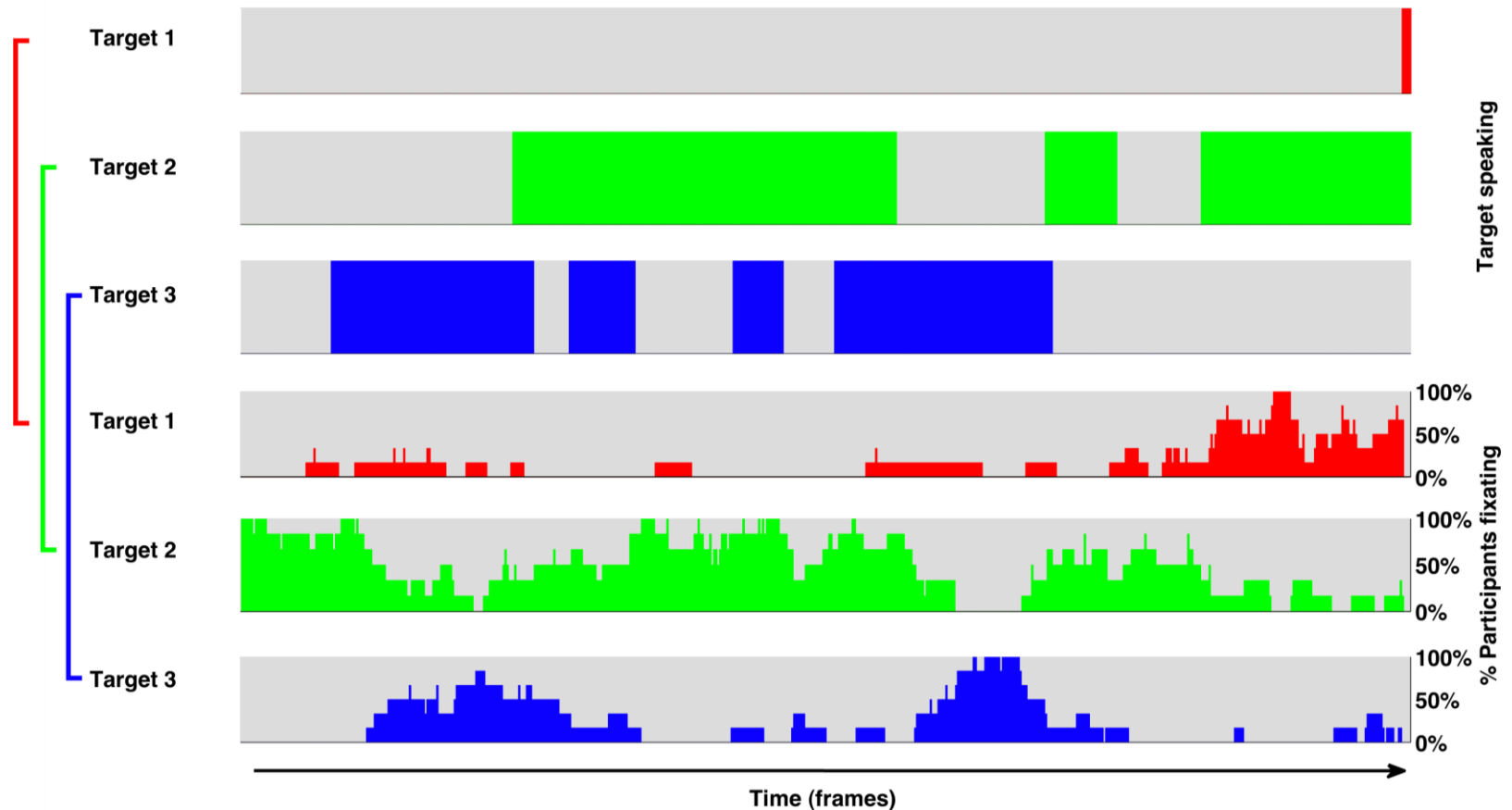and low status targets receiving the fewest fixations (all pairwise comparisons

$p<.05$).

**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. At this point we were interested in how the destination for each fixation was planned or guided, so we categorized 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. We analysed all fixations, and made a comparison between the proportion of these fixations that were directed at the person talking (at the start of that fixation) 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, $\eta_p^2$=.75). 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 ($F(2,48)=43.3$, $p<.001$, $\eta_p^2=.64$), 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, $\eta_p^2$=.02) 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, $\eta_p^2$=.41), 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 (see Richardson & Dale, 2005, for a similar approach). 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.  The pattern for gaze to precede speech was found across targets of different status, although the extent of the offset differed slightly (median lags for high, medium and low status = 3, 5 and 7 frames, respectively).  This variability would be interesting to study further, but, because it was not the focus of the present study, we included a relatively small number of clips and targets of each level of status and relatively few frames where low status targets were talking, so this potential status difference should be interpreted with caution.

**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).
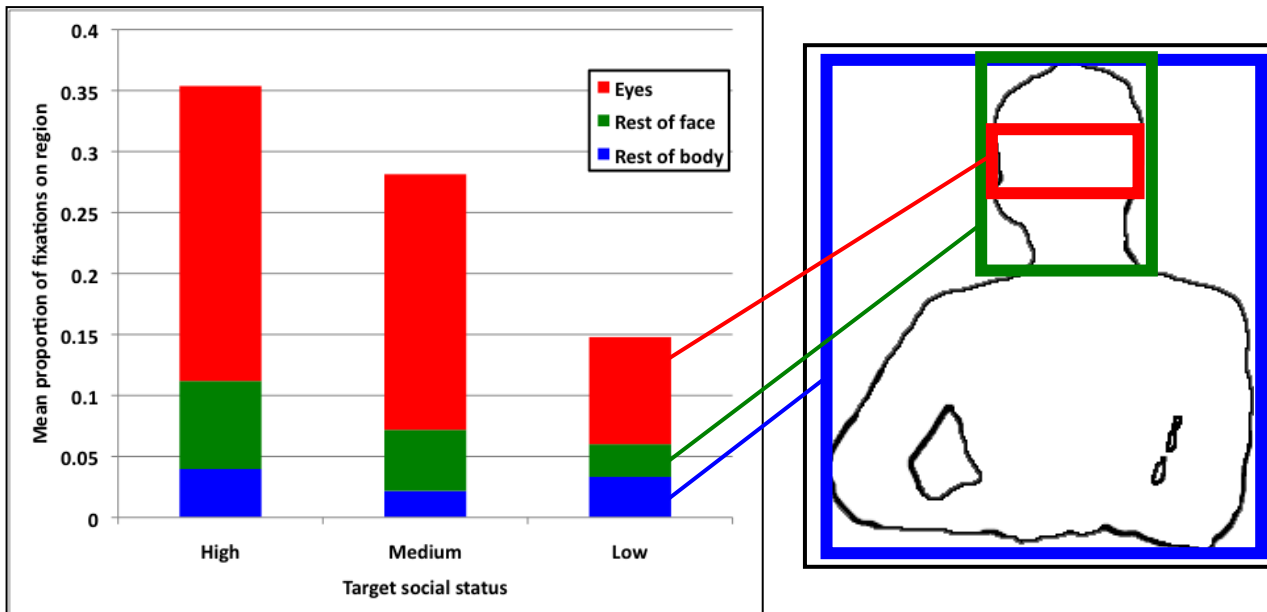
**Figure 4.** Measuring the amount of gaze given to different parts of the people was accomplished by defining moving areas of interest for the eyes and head (the relative sizes of which are depicted with a diagram of one target in the right panel). The proportion of fixations on each of these regions, 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 ($F(2,48)=31.8$, $p<.001$, $\eta_p^2=.57$). There was also a main effect of region of interest ($F(2,48)=74.7$, $p<.001$, $\eta_p^2=.78$). 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, $\eta_p^2$=.48), 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, $\eta_p^2$=.82) and medium-status targets (F(2,23)=77.3, p<.001, $\eta_p^2$=.87). 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, $\eta_p^2$=.73). 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).

Humans have physiologically evolved to communicate their eye gaze direction to others (for example by having a high contrast between the iris and sclera see Kobayashi & Koshima, 1997), but the function of attentional orienting to the eyes is not fully understood. Evidence suggests, however, that the ability to share others' goals and intentions (i.e., theory of mind) is fundamentally linked to orienting to another's gaze (Tomasello, Carpenter, Call, Behne, & Moll, 2005). For example, evidence from individuals with autism (Baron-Cohen, 1995), psychopathy (Dadds, El Masry, Wimalaweera, & Guastella, 2008), and patients with selective

damage to the amygdala (Adolphs, et al., 2005), confirms that deficits in processing emotion and theory of mind are often accompanied by a reduced tendency to look at the eyes. The hormone oxytocin enhances both social approach behaviour and fixations to the eyes in humans, confirming this link and suggesting part of the mechanism (Guastella, Mitchell, & Dadds, 2008). In our naturalistic task, watching a social interaction while thinking about some of the people involved, the eyes were spontaneously selected by observers as being especially useful. This is consistent with a theoretical model that attentional processing of the eyes functions to enhance the perception of the target's external goals (e.g. who they are talking to) as well as their internal emotions, intentions and beliefs (e.g. how they feel about that person; see Birmingham & Kingstone, 2009 for a recent review).

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.  Our control study demonstrated that the selection

of high status individuals occurred even when we asked observers to think about

people that they would not want to work with, demonstrating that the effect of

status on gaze direction was not specific to our instructions, but rather represents a

spontaneously adopted pattern.

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, and from other group members' behaviors and responses toward them (e.g., others asked them for advice, deferred to their opinions, etc.), which jointly determine a target's position in the social hierarchy. In particular, status rank differentiation may be the result of individual differences in dominance (i.e., an individual's ability to be forceful and intimidate others) and prestige (i.e., an individual's ability to demonstrate valued skills and expertise; Henrich & Gil-White, 2001). Indeed, in the group task described here, individuals who were perceived to be prestigious or dominant by other group members attained the highest level of overall status and influence over others (Cheng et al., in prep). Further research is necessary to identify the specific behaviours (e.g., elements of nonverbal or verbal behavior) that are in themselves salient attractors of attention, and to address the question of whether a conscious attribution of status is necessary for the effect on attention. 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.

Regardless of these issues, the present findings suggest that observers can very quickly ascertain who the high-status members of a group are, and 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; Cheng et al., in prep). 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 (i.e., if they demonstrate prestige), and monitor potential threats or attacks from these more powerful conspecifics (i.e., those who demonstrate dominance). Converging lines of research show that young children (3-4 year olds) automatically track the gaze of other people, and preferentially imitate the preferences of those most gazed at by others, even when they are alone (Chudek, Heller, Birch & Henrich, submitted). 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).

Given that we interpret the differences in eye movements as evidence for sensitivity to the social context, and particularly the status hierarchy, it is likely that these effects will be moderated by individual differences known to influence this sensitivity. For example, observers who are themselves considered low social status should be more inclined to look at high social status individuals, so as to monitor and learn from their superiors (in the same sense that monkeys only follow the gaze of conspecifics who are of higher social status; Shepherd, et al., 2006). In fact, previous research suggests that observers low in feelings of belongingness show heightened gaze-following tendencies (Wilkowski, Robinson, & Friesen, 2009). The gender of observers and targets may also make a difference as women and men prioritize different attributes in potential mates (Li, Bailey, Kenrick, & Linsenmeier,

2002).  Such predictions are fuel for further research, and the paradigm used here may prove fruitful for testing them.

Another factor that might affect results is targets' relative attractiveness. Attractive people are thought to draw attention (Maner, et al., 2003), and attractiveness  has been found to predict high social status in some groups (Anderson, John, Keltner, & Kring, 2001).  The present study design may provide a suitable context for testing the effects of attractiveness on attention in dynamic contexts, particularly as there is evidence that attractiveness inferred from watching dynamic video might prove to be different from image-based, "physical" attractiveness (Riggio, Widamen, Tucker, & Salinas, 1991).  While we did not obtain ratings of attractiveness in the present study, we did examine the related construct of interpersonal liking, and found that: (a) the effects of social status on attention remained significant even after controlling for ratings of peer liking, and (b) although liking had a small positive effect on attention, this effect was reduced to non-significance once speaking time was accounted for. This apparent null association between liking and attention leads to the speculation that physical attractiveness is not likely to completely explain the observed attentional biases, which clearly point to a status-advantage. Nonetheless, disentangling these influences on attention would be a fruitful avenue for further research. However, it is important to bear in mind that status and attractiveness may be too interdependent to fully tease apart these relations; given evidence that attractiveness partly determines status, controlling for this variable may be a case of "throwing out the baby with the bathwater".

By examining 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). Intriguingly, deaf people also tend to look at the eyes of others communicating with sign language, although beginners also look at the mouth (Emmorey, Thompson, & Colvin, 2009). The vast majority of these studies displayed the face of a single speaker performing a monologue, but we can extend the importance of gaze to the eyes to more realistic, three-party conversations.

In research involving 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 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.  It is also interesting to consider the pattern of gaze preceding speech in the context of research into eye movements in speech perception.  When participants listen to spoken sentences they recognize objects named in the sentence and move their eyes to these objects a few hundred milliseconds *after* the beginning of the word (Allopenna, Magnuson, & Tanenhaus, 1998).  In their study of gaze coordination, Richardson and Dale (2005) found that when viewing a picture of characters from a TV show, a listener's eye movements were best aligned with those of the speaker after a 2 second delay.  In these studies, presumably, participants take time to process what they hear and plan appropriate eye movements.  Why did we find that gaze *preceded* the speaker?  One reason may be the dynamic and social nature of our targets.  The targets were talking about abstract items, and we found a temporal synchrony with fixations to the speaker, rather than to any explicitly mentioned object.  In fact, the pattern we observe is more similar to that found in studies of speech *production*, where people describing a scene tend to look at

objects up to a second prior to naming them (Griffin & Bock, 2000). Similarly, if the pragmatic or linguistic context of speech is predictive of what will be mentioned next, participants make anticipatory eye movements that precede the utterance (Altmann & Kamide, 2009), something that also occurs in realistic interactions with a partner (Brown-Schmidt & Tanenhaus, 2008). We suggest that the pattern we observed demonstrates the social context of the conversation to which our observers were attuned. 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.

**Acknowledgements**

support of a Commonwealth Fellowship from the Government of Canada to TF; and a Social Sciences Research Council of Canada award to JC (File #s 767-2009-2108).

# 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.

Adolphs, R., Gosselin, F., Buchanan, T. W., Tranel, D., Schyns, P., & Damasio, A. R. (2005). A mechanism for impaired fear recognition after amygdala damage. *Nature, 433*(7021), 68-72.

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language, 38*(4), 419-439.

Altmann, G. T. M., & Kamide, Y. (2009). Discourse-mediation of the mapping between language and the visual world: Eye movements and mental representation. *Cognition, 111*(1), 55-71.

Anderson, C., John, O. P., Keltner, D., & Kring, A. M. (2001). Who attains social status? Effects of personality and physical attractiveness in social groups. *Journal of Personality and Social Psychology*, *81*, 116-132.

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.

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

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

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

Brown-Schmidt, S., & Tanenhaus, M. K. (2008). Real-time investigation of referential domains in unscripted conversation: A targeted language game approach. *Cognitive Science, 32*(4), 643-684.

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

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., Foulsham, T. & Kingstone, A. *The emergence of dominance and prestige hierarchies in unacquainted social groups.* Manuscript in preparation.

Chudek, M., Heller, S., Birch, S., & Henrich, J. *Prestige-biased learning in children: Attention from others as a cue for cultural transmission.* Manuscript submitted for publication.

Dadds, M. R., El Masry, Y., Wimalaweera, S., & Guastella, A. J. (2008). Reduced eye gaze explains "Fear Blindness" in childhood psychopathic traits. *Journal of the American Academy of Child and Adolescent Psychiatry, 47*(4), 455-463.

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.

Emmorey, K., Thompson, R., & Colvin, R. (2009). Eye Gaze During Comprehension of American Sign Language by Native and Beginning Signers. *Journal of Deaf Studies and Deaf Education, 14*(2), 237-243.

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.

Griffin, Z. M., & Bock, K. (2000). What the eyes say about speaking. *Psychological Science, 11*(4), 274-279.

Guastella, A. J., Mitchell, P. B., & Dadds, M. R. (2008). Oxytocin increases gaze to the eye region of human faces. *Biological Psychiatry, 63*(1), 3-5.

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.

Kobayashi, H. & Koshima, S. (1997). Unique morphology of the human eye. *Nature*, 387, 767–768.

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.

Li, N. P., Bailey, J. M., Kenrick, D. T., & Linsenmeier, J. A. W. (2002). The necessities and luxuries of mate preferences: Testing the tradeoffs. *Journal of Personality and Social Psychology*, *82*, 947-955.

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.

Maner, J. K., Kenrick, D. T., Becker, D. V., Delton, A. W., Hofer, B., Wilbur, C. J., et al. (2003). Sexually selective cognition: Beauty captures the mind of the beholder. *Journal of Personality and Social Psychology, 85*(6), 1107-1120.

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. (2005). Looking to understand: The coupling between speakers' and listeners' eye movements and its relationship to discourse comprehension. *Cognitive Science, 29*(6), 1045-1060.

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.

Riggio, R. E., Widamen, K. F., Tucker, J. S., & Salinas, C. (1991). Beauty is more than skin deep: Components of attractiveness. *Basic and Applied Social Psychology, 12,* 423-439.

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.

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and Sharing Intentions: The Origins of Cultural Cognition. *Behavioral and Brain Sciences*, *28*, 675-691.

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.

Wilkowski, B. M., Robinson, M. D., & Friesen, C. K. (2009) Gaze-triggered orienting as a tool of the belongingness self-regulation system. *Psychological Science, 20*, 495–501.

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