Don't Look Now: The relationship between mutual gaze, task performance and staring in Second Life

Sara Dalzel-Job (S.Dalzel-Job@sms.ed.ac.uk)
University of Edinburgh, School of Informatics, Edinburgh, UK
Jon Oberlander (j.oberlander@ed.ac.uk)
University of Edinburgh, School of Informatics, Edinburgh, UK
Tim J Smith (tj.smith@bbk.ac.uk)
Birkbeck, University of London, London, UK

Abstract

Mutual gaze is important to social interaction, and can also facilitate task performance. Previous work has assumed that staring at someone maximises mutual gaze. Eye-tracking is used to explore this claim, along with the relationship between mutual gaze and task performance. Two participants – Instruction Giver (IG) and Instruction Follower (IF) – communicated via avatars in Second Life to solve simple arithmetic tasks. There were two conditions: *staring* (the IG's avatar stared continuously at the IF); and *not-staring*, (IG's avatar looked at IF and task-relevant objects). Instead of maximising mutual gaze, constant staring actually showed evidence of *decreasing* eye contact within the dyad. Mutual gaze was positively correlated with task performance scores, but only in the *not-staring* condition. When not engaged in mutual gaze, the IF looked more at task-related objects in the *not-staring* condition than in the *staring* condition. Implications and possible future work on social interaction are discussed.

Keywords: Mutual Gaze; Second Life; Task Performance; Staring; Joint Attention.

Introduction

Non-verbal communication is an important contributor to successful social interaction. Gaze direction, in particular, provides rich social information, such as social accessibility: *mutual gaze,* or eye contact, can indicate that a conversational channel is open, and that an interlocutor is willing to engage, or continue to engage, in an interaction. *Joint attention*, or being aware of a conversational partner's eye movements, and consequently focusing on the object of their attention, is a skill that is developed in infancy (Corkum & Moore, 1998) and is widely used during conversation. For example, we can infer the object a conversational partner is referring to by following their gaze. Mutual gaze has also been reported to facilitate performance on cognitive tasks. Early work by Fry and Smith (1975) found that increased eye contact resulted in better task performance on a digit encoding task. Fullwood and Doherty-Sneddon (2006) discovered that more looking by a confederate at the camera during a video presentation maximised the subsequent recall by the viewer.

If mutual gaze does, indeed, facilitate task performance, it would be pertinent to find out how to maximise the amount of mutual gaze between a conversational pair (*dyad*). Fry and Smith (1975) merely state that "Eye contact was manipulated" during the experiment, with an Instruction giver giving her conversational partner either "as much eye contact as possible" or "as little eye contact as possible", depending on the condition (p2). One of the fundamental aspects of mutual gaze is that it is a joint action – one cannot independently engage in mutual gaze, and therefore cannot *give* (or *be given*) eye contact, as it is an inherently mutual activity. In a contemporary adaptation of Argyle and Dean's (1965) exploration of Equilibrium Theory, Bailenson et al. (2001) investigated the amount of interpersonal distance an individual maintained from a virtual being in an immersive virtual environment. This virtual being was programmed so that it "engaged them [the participant] in eye contact (that is, mutual gaze behavior)" (p1). This, again, implies that a single person has the ability to independently control the amount of mutual gaze that occurs between themself and a conversational partner. It assumes that all one must do to in order to engage in a maximum amount of mutual gaze is to stare at a conversational partner. It could, however, be seen as socially inappropriate to stare constantly at someone, since "To be subjected to the continual gaze of another is a very unnerving experience, for to be the object of another's attention is to be vulnerable to him." (Kendon, 1967, p48). Consequently, it is entirely possible that being constantly stared at could actually *reduce* one's willingness to engage in mutual gaze, rather than maximise it.

It is further possible that constant staring may be in some way detrimental to task performance. If, as Kendon suggests, being stared at is unnerving, then it may be that, during a task-based interaction, a stared-at party (as opposed to the *starer*) will deflect their gaze to *anywhere other than the eyes of the person who is staring at them*, rather than directing it towards a functional object that could assist in the completion of the task at hand. It is therefore also of interest to establish where the stared-at party is looking when not engaging in mutual gaze, and how this looking behaviour differs if not being stared at. Is it task-focused looking, or instead *anywhere but at the starer*?

To investigate these issues, a suitable platform is required. For one conversational partner to stare continuously at the other, a high level of control over one of the interlocutors' eye movements is needed, since this is not generally a natural human behaviour. It is also necessary for the eye movements of the non-staring partner to be recorded during the interaction, along with the task performance scores, thus addressing the questions of whether constant staring by one individual at another will maximise the overall amount of mutual gaze between the dyad, and how the overall amount of mutual gaze relates to the subsequent performance.

Gaze in Second Life

With the increasing interest in virtual environments (VEs) over recent years, and along with their rapid development, has come an understanding of the benefits of using such platforms for the study of social interaction.

Second Life (SL) is a 3D virtual environment in which users are able to interact with other users and agents via an avatar (see http://secondlife.com/). The default (and manipulated) avatar eye and body movements are very human-like, enabling the experimenter to draw inferences from interactions with avatars/agents and potentially apply them to human-human interactions. Bailenson et al. (2001) found that, in terms of inter-personal distance, people treated agents similarly to the way they treat real humans. The interface is relatively easy to use, and scripting facilities allow the import of a given task or paradigm, such as a problem to be jointly solved by two people, mirroring a real-world interaction in a more controllable environment. This paradigm can then be easily adapted to different domains. An online (as opposed to post-test) evaluation can be made of how individuals respond to a task by capturing the screen during the interaction, superimposing gaze behaviour, and analysing it in conjunction with other dependent variables, such as task performance. It is possible to access SL on the three main computer platforms. Given all of this, it constitutes a useful means for studying social interaction within a controlled environment.

Much of the previous research into eye movements in SL has been dedicated to using eyes to control a user's avatar, a method especially valuable for individuals with disabilities that inhibit them from using a standard mouse and keyboard (e.g. Vickers et al., 2008). Dalzel-Job, Nicol and Oberlander (2008), however, recorded users' eye movements during a task-orientated interaction with a programmed avatar (agent) to investigate how individuals respond to informative compared with redundant gestures in SL. Yee et al. (2007) investigated mutual gaze and interpersonal distance with an avatar in a virtual environment, and found that, on the whole, such interactions were governed by the same social norms as those in the real world. This was another variation of Argyle's Equilibrium Theory paradigm, although they were observing eye contact and interpersonal distance between avatars in SL, rather than the people controlling them. This would probably not give an accurate indication as to the eye movements of the users driving the avatar; it would only indicate that their avatars were making eye contact. There have been no studies devoted to the measurement of a user's eye movements during interaction with another user in SL.

The first question of interest is does constant staring by one conversational partner at another maximise the amount of mutual gaze between the dyad, as previously predicted by Bailenson et al. (2001) and Fry and Smith (1975)?

Secondly, does mutual gaze facilitate task performance, as found by Fullwood and Doherty-Sneddon (2006) and Fry and Smith (1975)? What, if any, is the relationship between the amount of mutual gaze and the task performance, and how does this relationship change when one conversational partner stares continuously at the other?

Finally, if, as predicted, staring does *not* maximise mutual gaze, then where is the stared-at party looking when not returning mutual gaze? How do these eye movements vary when being stared at compared to not being stared at? It is predicted that being stared at will increase the likelihood of the stared-at party looking at task-irrelevant objects (when not engaging in mutual gaze), but when not being stared at, they will be more likely to look at task-relevant objects (when not returning mutual gaze).

Method

Dyads (two participants) – an Instruction Giver (IG) and an Instruction Follower (IF) – completed relatively simple arithmetic problems (such as 8+3+2; see Instruction Tiles in figure 2) under two conditions – *staring* and *not-staring*. In the staring condition, the IG's avatar stared continuously at the IF during the interaction, and in the not-staring condition, the IG's avatar looked at the IF intermittently, during the interaction. The participants were fully aware that they were interacting with another human being.

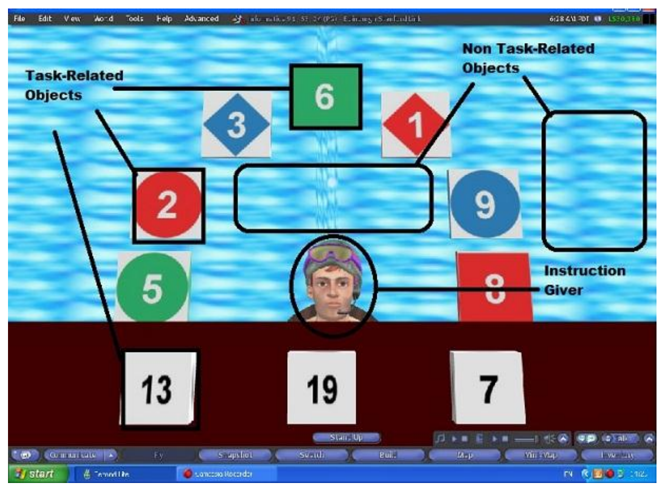

Figure 1: Instruction Follower's View and Regions of Interest (black outlines; not visible in experiment)

The first dependent measure was task performance, as measured by how many of 15 tasks the IF correctly completed under each condition. The second dependent measure was the proportion of the interaction during which the IF looked at pre-defined regions on the screen. The IF's screen was divided into 3 regions of interest: the IG's avatar (*IG)*, the tiles (*task-related objects*) and anything else (*non task-related objects*) (see Figure 1).

Participants

52 participants (mean age 23.4; 27F) were randomly assigned to pairs and were tested for colour-blindness prior to the procedure using the PseudoIsochromatic Plate Ishihara Compatible (PIP) Color Vision Test 24 Plate Edition (see http://colorvisiontesting.com/ishihara.htm). 6 dyads were excluded from analysis because of synchronisation issues between the audio and video recorded during the experiment.

Apparatus

Participants viewed the experiment on a 19 inch CRT display. The IF used a standard mouse to respond to questions asked. In the staring condition, the IG was instructed not to touch the mouse or keyboard, and in the not-staring condition was told he should move the mouse to hover the cursor over the tile that he was describing, which resulted in the IG's avatar looking at the tiles that were being described. An SR-Research EyeLink II head-mounted tracking system was used during the study to record eye movements of both participants. The sample rate was set at 500Hz and the participants' dominant eye was tracked monocularly. Only the IF's eye behaviour is reported in the current paper. Additionally, the IF wore a set of headphones and the IG wore a microphone headset to enable the follower to hear the IG's instructions via his avatar in SL. A 9-point calibration matrix was used at the start of each participant's experiment and between blocks if required. Camtasia Studio (TechSmith Ltd) recorded what each participant could see on the screen throughout the procedure, along with audio (the IG's instructions) throughout the experiment, generating movie files for analysis in conjunction with the eye movements.

Stimuli

A building comprising of 1 large closed room was built on VUE, the University of Edinburgh's Island within SL (see http://www.vue.ed.ac.uk). There were 2 chairs facing each other within the room with a glass screen between them. The participants' avatars sat on the chairs. In front of each was a panel that was hidden from the other participant's view. On the IG's side the panel contained Instruction Tiles, the contents of which were to be conveyed to the IF (Figure 2). On the IF's side were 3 Answer Tiles on which were presented 3 multiple-choice answers (Figure 1). On the glass screen between the 2 avatars were 7 Stimulus Tiles, which were visible to both participants. Each Stimulus Tile had a number on a background of a shape of a particular colour (Figure 1; Figure 2).

The users' view was pre-programmed so that they were 'seeing' through their avatar's eyes, resulting in opportunities for mutual gaze.

The IG conveyed each of 2 blocks of the 15 arithmetic problems verbally to the IF via their avatars in SL. The two sets of tasks were counterbalanced between experimental conditions.

All of the tiles – the instruction, stimulus and response tiles - were created within SL and textures were attached as required throughout the experiment. All stimuli materials were created using Microsoft Paint version 5.1 and GIMP (GNU Image Manipulation Program).

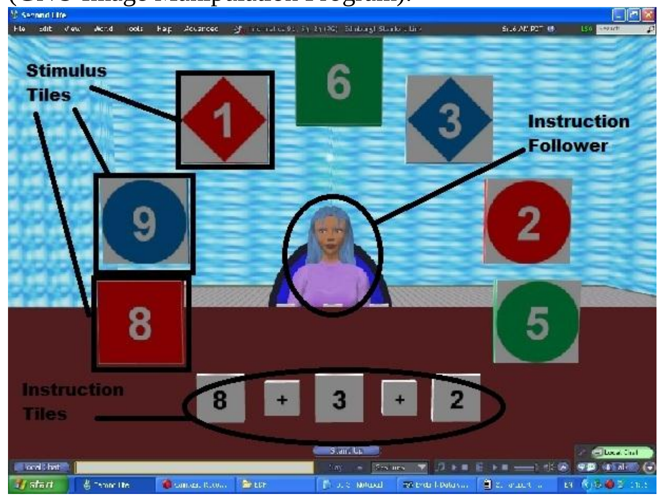

Figure 2: Instruction Giver's View and Regions of Interest (black outlines; not visible in experiment)

Design

In a within-subjects design, all participants carried out the 15 tasks under each of the 2 conditions – *staring* and *not-staring***.** The tasks were counterbalanced between the participants for the 2 conditions to control for effects due to task itself. The 2 conditions were as follows:

1. The IG's avatar looks directly at the IF, providing a *staring* condition. This was achieved by asking the IG not to move the mouse, resulting in the default behaviour of an avatar in SL – staring straight ahead – i.e. at the follower.
2. The IG's avatar looks at the tiles while describing them, and looks at the IF for the remaining duration, providing a *not-staring* condition. This was achieved by asking the IG to move the cursor so that it hovered over the tile they were describing. This automatically moves the IG's avatar's gaze to the focused tile, and then returns to the default 'looking-straight-ahead' (i.e. at the IF) after a few seconds. Under this condition, the gaze of the IG is informative – his avatar looks at the tile he is describing – but it must be noted that this visual information is redundant, since the IF gets all of the details required to complete the task verbally.

The IF was unaware of the IG's instructions to manipulate the gaze of his avatar.

The order of the conditions remained constant for each dyad to reduce potential for errors made by the IG; since they were only required to manipulate the gaze of their avatar in the not-staring condition, the instructions to move the mouse were only given after the conclusion of the staring condition, thus reducing any accidental mouse moving during that condition. Each task was presented to the IG via the Instruction Tiles, for him to convey it to the IF. In both conditions, the IG was allowed to formulate the instructions as he or she wished, as long as the numbers were not mentioned. In the IG's view in Figure 2, for example, the IG would say 'red square plus blue diamond plus green circle'. The IF selected the correct answer by clicking on one of the three answer tiles, so the correct action here would be to select the left-most response tile, indicating that 13 was the correct answer (see Figure 1). This resulted in the texture on the tiles being updated for the next task.

Since the comparison to be made was between the 2 conditions – i.e. a related design (within subjects) – and it could be assumed that the style of instructions was consistent throughout the experiment, a comparison between blocks subtract out individual differences in instructions.

Results

Mutual Gaze

It was anticipated that constant staring by one conversational partner at another will *not* maximise the amount of mutual gaze between the dyad. An initial analysis looked at the proportion of the trial that the IF spent looking at the IG's avatar in the staring and the not-staring conditions, asking: was there a difference between the amount of attention that the avatar attracted in the staring and the not-staring conditions?

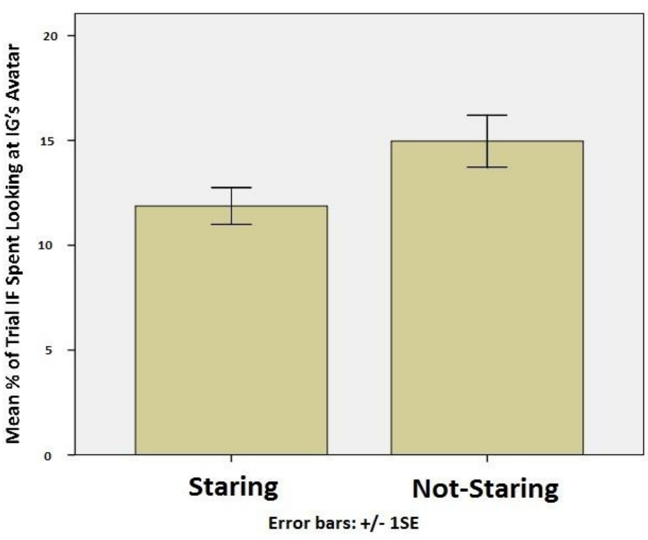

Figure 3: Mean % of Trial IF Spent Looking at IG's Avatar in Staring and Not-Staring Conditions

A paired samples t-test found that the IF spent significantly more time looking at the IG's avatar in the not-staring than in the staring condition (M=14.96, SD=5.81 and M=11.87, SD=4.12, respectively respectively), $t(21)=2.705$; $p<.05$ (see Figure 3).

To investigate the amount of mutual gaze that the dyad engaged in under each condition, the absolute amounts of mutual gaze were compared with a paired samples t-test. Although approaching significance, there was found to be no overall difference between the conditions ($p>.05$, NS). To investigate this further, the proportion of the total number of opportunities for mutual gaze that were taken up by the IF was compared for the staring and not-staring conditions (see Figure 4). The total opportunities for mutual gaze equated to all of the times when the IG was looking at the IF. When the IF looked back at the IG, these opportunities were said to be taken up. In the staring condition, this uptake was the same as % of the trial during which the IF looked at the IG (as in Figure 3). It was found that there were significantly more opportunities for mutual gaze taken up in the not-staring condition than in the staring condition (M=18.08, SD=4.12; M=11.87, SD=10.13 respectively), $t(21)=3.417$; $p<.005$.

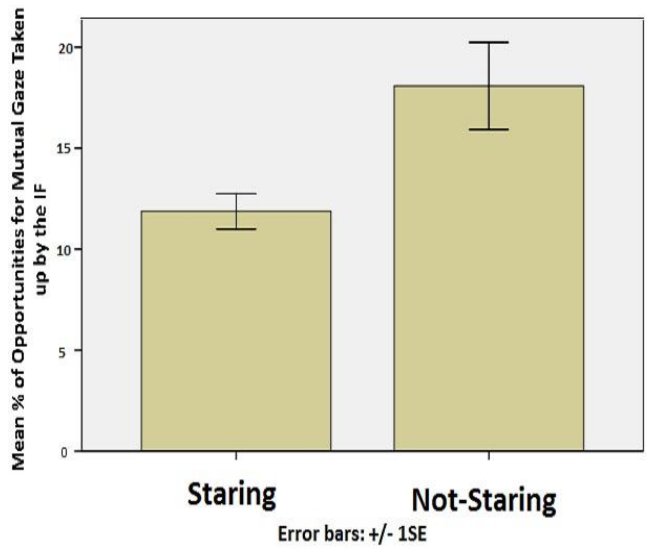

Figure 4: Mean % of Opportunities for Mutual Gaze Taken up by IF

Task Performance

It was expected that there would be a positive correlation between the proportion of mutual gaze between the dyad and task performance score (measured by how many tasks out of 15 were completed correctly) in both conditions.

Before analysis of the task performance scores, three of the dyads had to be excluded, since they had failed to understand the instructions, and therefore responded to the questions incorrectly. The remaining 18 dyads' task performance scores were compared. A Wilcoxon Signed Ranks Test found there to be no significant difference between the overall task performance scores in the staring and not-staring conditions ($Z = -.303$, $p=.71$). Indeed, median task performance scores were 14 in both conditions.

A Spearman's rho correlation found that in the not-staring condition, task performance was significantly correlated with the proportion of trial spent in mutual gaze ($r_s = .48$ (18), $p< .05$). In the staring condition, however, it was found that, despite a positive trend, there was no significant relationship between task performance and mutual gaze (r_s = .36 (18), *p*=.062).

Since there was no overall difference between the task performance scores in the staring and not-staring conditions, it was of interest to investigate why mutual gaze had a differing effect on task performance in the two conditions. The IF's eye movements during missed opportunities for mutual gaze were compared under the two independent variables. This comprised all of the occasions when IF did not look at IG in the staring condition, compared with all the times in the not-staring condition when IG is looking at IF, but IF is looking. The distribution of IF looking behaviour during all such opportunities for the staring and not-staring conditions, can be seen in figures 5 and 6, respectively. In the not-staring condition, this time comprised approximately 27% of the total trial.

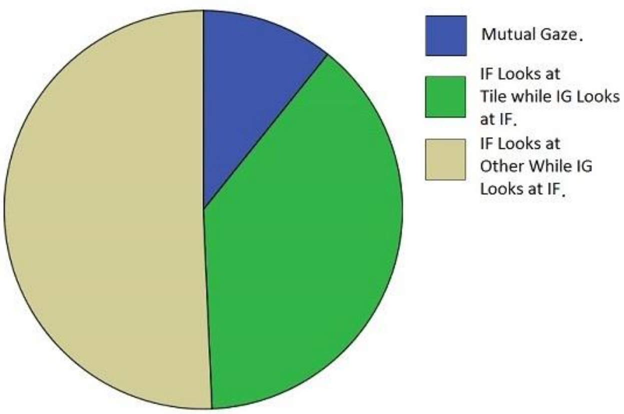

Figure 5: Mean % of Staring Trial that IF Spent in Each Looking Behaviour

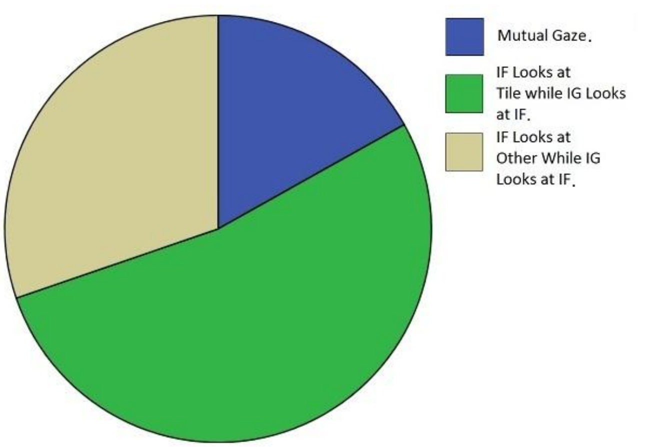

Figure 6: Mean % of Not-Staring Trial Spent in Each Looking Behaviour: time during which IG is looking at IF

It was predicted that the IF would look anywhere apart from at the IG when being stared at, rather than at task-orientated stimuli; followers would spend a larger proportion of the trial looking at task-irrelevant objects ('other') in the staring condition than in the not-staring condition. This difference was found to be significant. The ratio of the proportion of the trial that the IF spends looking at non-task-related or 'other' compared with task-related, or 'tiles', was found to be significantly higher in the staring condition than in the not-staring condition ($t(19)=3.509$; $p<.01$).

Discussion

We were interested here in whether constant staring by one conversational partner at another maximises the amount of mutual gaze between the dyad. It was found that if an Instruction Follower is being stared at, he is likely to spend less time looking at the face of the person staring – the Instruction Giver. It was also found that, contrary to previous assumptions, having one conversational partner stare constantly at the other does not maximise the amount of mutual gaze between the dyad: there was no significant difference between the absolute amounts of mutual gaze in the staring and not-staring conditions. The IF had the opportunity to engage in mutual gaze at any time during the interaction in the staring condition, but there were fewer opportunities for mutual gaze in the other condition (approximately 27% of the not-staring, vs. 100% of the staring trial). There were, however, no more overall occurrences of mutual gaze in the staring condition than in the not-staring condition, despite the greater opportunities. In fact, a higher proportion of opportunities for mutual gaze were taken up in the not-staring condition than in the staring condition. So, far from maximising mutual gaze, staring resulted in a lower uptake of opportunities for mutual gaze: staring actually *decreases* mutual gaze.

It seems entirely reasonable to assume that there are social factors at work here, which discourage an individual from returning the stare of their conversational partner, to avoid being, as Kendon suggests, "vulnerable to him". It could be argued, however, that the IF looked more at the IG during the not-staring condition because of visual information that could assist in the completion of the task in this condition. Although this information is strictly redundant, this possible explanation will be tested in a further study with an additional baseline condition where the IG still looks at the tiles redundantly, but does not look at the IF during the procedure. Comparison between the conditions will help distinguish attention attracted for task-related reasons (i.e. because the IG is looking at the tiles) from that attracted for social reasons (i.e. because the IF wishes to engage in eye contact).

As predicted, the more mutual gaze there was between a dyad, the better the task performance. This only held true, however, when the IF was not being stared at. This suggests that if you want your interlocutor to retain the information that you are imparting, then you should try and maximise the amount of mutual gaze between the pair of you. But this does not involve staring: staring will not influence task performance in the same way that not staring can; staring maximises neither mutual gaze nor task performance. In future analysis, we will systematically explore the relationship between varying amounts of gaze by the IG and its effects on mutual gaze and task performance.

The finding that the IFs were less likely to spend their non-mutual-gaze periods looking at task-related objects in the staring condition than in the not-staring condition may go some way towards explaining the lack of relationship between mutual gaze and task performance in the staring condition. Directing gaze towards irrelevant objects does not help task performance.

In this study, the participants were fully aware that they were interacting with another human, the avatar behaviour was human-like and there is precedent for using virtual humans to investigate human-human interaction (Yee et al., 2007; Bailenson et al., 2001). But at this point, strong conclusions about face-to-face human-human behaviour cannot be drawn. The dependent variable *agency* will be included in the next study, meaning that users will either be told they are interacting with an avatar (human controlled) or an agent (computer controlled). This should foreground the differences between how people treat humans and computers within this paradigm. There is also scope for analogous face-to-face human-human experiments, to further test the relationship between human-avatar interaction and interaction in the real world.

Conclusions

The discovery that task performance can be facilitated by increasing mutual gaze has implications for many areas of life, from business meetings to pedagogy, including virtual teaching agents, and perhaps even face-to-face teaching. Mutual gaze matters during social interaction.

Further investigation should be made to establish how much looking by one conversational partner at another is optimal for mutual gaze and task performance on a given task. If mutual gaze can be optimised, then it follows that task performance may also be optimised. It is anticipated that this will take the form of a human-agent experiment within Second Life. Analysis will be made of IG's gaze behaviour from the current experiment, on which the eye movements of the agent in the next experiment will be based. Additional control conditions will be in place to help eliminate other possible explanations for variation in looking behaviours. It is anticipated that a face-to-face human-human experiment will validate these results, enabling the generalisation of future research using this paradigm to face-to-face interactions.

It would also be of interest to discover what is underlying the varying amount of mutual gaze that an individual is willing to engage in. In computer mediated communication, compared with face-to-face interactions, participants will experience an altered perception of the level of social accessibility of their interlocutor. When someone is staring at you, for example, do you perceive them to be more or less socially accessible, and how does this relate to your overall social perception of that individual? By looking into these factors, it should be possible to develop a more rounded model of mutual gaze, task performance, and the socio-cognitive factors underlying the two.

References

Argyle, M. & Dean, J. (1965). Eye-contact, distance and affiliation. *Sociometry 28(3), 289-304.* Sept 1965.

Bailenson, J., Blascovich, J., Beall, A. & Loomis, J. (2001). Equilibrium theory revisited: Mutual gaze and personal space in virtual environments, *Presence 10(6),* December 2001, 583-598.

Corkum, V. & Moore, C. (1998). The origins of joint visual attention in infants. *Developmental Psychology 34(1), 28-38.*

Dalzel-Job, S., Nicol, C. & Oberlander, J. (2008). Comparing behavioural and self-report measures of engagement with an embodied conversational agent: A first report on eye tracking in Second Life. In *Proceedings of the 2008 Symposium on Eye Tracking Research & Applications, Savannah, GA, March 26-28 2008.*

Fullwood, C. & Doherty-Sneddon, G. (2006). Effect of gazing at the camera during a video link on recall. *Applied Ergonomics 37; 167–175.*

Fry, R. & Smith, G.F. (1975). The effects of feedback and eye contact on performance of a digit-encoding task. *Journal of Social Psychology 96(1): 145–146.*

Kendon, A. (1967). Some functions of gaze-direction in social interaction. *Acta Psychologica 26;* 22-63.

Vickers, S., Istance, H.O., Hyrskykari, A., Ali, N. & Bates, R. (2008). Keeping an eye on the game: Eye gaze interaction with massively multiplayer online games and virtual communities for motor impaired users. *Proceedings of the 7th International Conference on Disability, Virtual Reality and Associated Technologies.*

Yee, N., Bailenson, J.N., Urbanek, M., Chang, F. & Merget, D. (2007). The unbearable likeness of being digital: the persistence of nonverbal social norms in online virtual environments. *Cyberpsychol. Behav. 10; 115-121.*

Acknowledgments

This work was supported by the ESRC and Edinburgh's Informatics Graduate School. Thanks also to the JAST, Indigo and JAMES projects for support for the overall programme. Input into this project by Jeffrey Dalton of AIAI, University of Edinburgh is gratefully acknowledged. Thanks also to the reviewers for their constructive critique.