Visual attention during conversation: an investigation using real-world stimuli.

Miss Jessica Dawson
BSc Psychology
MSc Research Methods in Psychology
AFHEA

Supervisor: Dr Tom Foulsham

A thesis submitted for the degree of Doctor of Philosophy

Department of Psychology
University of Essex
Submitted in September 2021

This thesis is supported by an ESRC SeNSS Scholarship

COVID-19 Impact Statement

Due to the unprecedented global pandemic, it should be noted that my research for this thesis has been slightly adjusted to ensure end goals are met. One particular aspect I wish to note, is a mobile eye tracking study which I present the planned details of under Appendix 5. I had gained ethical approval and was about to begin pilot testing when our offices were shut down and hence, I was unable to collect data for this experiment. The research would have offered an additional exploration of the effect of eye contact and gaze aversion in a live setting. I feel the research is still very relevant and has importance in the research area and therefore, I hope to explore this in the future when restrictions allow.

In response to the pandemic and the lack of in person testing, I have taken the initiative to run an additional three, initially unplanned experiments online (within Chapter 2). These experiments offer a slight tangent to my main research question, looking at the method of using mobile eye tracking data and explores the subjectivity I experienced myself during data analysis. This adds an additional methodological consideration, which previously would not have been explored in this thesis. Hence, I strongly believe, this strengthens the thesis as a whole and more importantly offers empirical knowledge about methods of analysing mobile eye tracking data which is relevant to all researchers which use this technique.

Acknowledgments

First, I would like to thank SeNSS (ESRC) for funding this research. I feel very privileged that they believed in me and made this research possible. I am also thankful to Dr Gustav Kuhn and Dr Eva Gutierrez-Sigut who have taken time to examine my thesis.

My PhD Supervisor Tom Foulsham has been an exceptional mentor throughout my four years. His wisdom and generosity with his time has meant I have learnt far more than I thought was possible. I will be forever grateful.

A big thank you also to Alan Kingstone and the University of British Columbia for allowing me to visit and learn from you.

On a more personal note, thank you to my Mum and George for putting up with my tears and believing in me when I didn't believe in myself. My Mum has made sacrifices to always put me and my education first. I know you both don't understand this thesis but thank you for persevering with me! Also, a big thank you to Auntie Bev for the relentless proof-reading, you're a star!

Finally, I would like to dedicate this thesis to my late grandparents for drilling my times tables and spellings into me from a young age – I hope I did you proud.

Contents

Thesis Abstract

This research investigates how people visually attend to each other in realistic settings. In particular, I explore how people move their eyes to attend to speakers during social situations. I examine which signalling cues are crucial to social interactions and how they work in conjunction to enable successful conversation in humans. Furthermore, a main aim of this research is to explore eye movement when participants are live or are third-party observers. Overall, using a range of techniques, the research has demonstrated the benefit of using both audio and visual cues to guide conversation following; how viewing the eyes of the speakers and their spatial location facilitates this; as well as an investigation of social attention in those with traits of disorders. Moreover, a key finding of the thesis is demonstrating the similarities in live eye-movements and third-party observations. Overall, the thesis offers a comprehensive account of which factors attract visual attention to speakers and facilitate conversation following.

Lay Introduction

Imagine you are engaged in conversation during a meeting at work. Sat around a table with your colleagues, you alternate who is speaking and who is listening, aware of when it is socially appropriate to speak and to listen. Your head movement, gestures, and tone of voice all signal when it is time for the alternation, but so do your eyes. Without realising, the eyes of you and your colleagues engage in a sophisticated pattern which allows you to not only gain information but also signal to your colleagues when you would like a response. This complex arrangement of wandering eyes is vital for comprehensive social engagement.

It may seem like an obvious observation that during a social interaction we look at the person who is speaking. Previous studies have suggested this is a social trait and an innate behaviour observed from infancy (Farroni et al., 2002). It could be argued that we look to the person that is speaking in order to gain information and help us to comprehend the social situation. However, what is it specifically that initially grabs and captivates our attention in a speaker, and what do we gain from looking at them? Research in various areas of psychology has demonstrated that the eyes of others are a dominant feature that we tend to fixate upon in social interactions. The eyes are not only used to gain visual information from the world around us, but our eyes can also act as a means to signal (Jarick & Kingstone, 2015). As humans we partake in a complex eye movement pattern where we can use our gaze to signal to others when it is their turn to speak. On attempting to measure the effect of eye gaze when engaging in a natural conversation, eyes are difficult to isolate on their own. Using 'real-world' stimuli, this thesis explores what determines where our visual attention is directed during conversation in both live and video settings.

Thesis flow

This thesis begins with a literature review of visual attention studies in terms of their history, classic and prominent studies in the field and what we know about visual attention to social stimuli and conversation thus far (Chapter 1). Chapter 2 offers an evaluation of the methods used to study eye movement, inclusive of three experiments, which explore considerations when coding eye-tracking data and theory of mind. Chapter 3 includes an explorative study comparing group size and eye contact in a live setting. Audiovisual cues are manipulated to assess their effect on visual attention in Chapter 4, with two experiments exploring this further. In the final experimental chapter, I examine how atypical differences (more specifically traits of ASD and ADHD) affect eye-movement to group conversation clips with and without the eyes of targets as an available cue. Finally, I close summarising overall findings and potential future directions.

Authors Notes

The following points should be noted. Elements of Experiment 1, 2 and 3 (Chapter 2) are published in Scientific Reports (with collaborators Alan Kingstone and Tom Foulsham). This work has since been restructured for this thesis. Thus, some of this content within this thesis replicates this publication. The explorative data for Experiment 4 (Chapter 3), was collected with a colleague, Jonas Großekathöfer. Experiment 5 (Chapter 4) is published as it stands in a special issue of Visual Cognition (with Tom Foulsham as a collaborator). This is reiterated throughout the thesis as well as explicitly stating where further collaborations have been made.

As someone who has worked closely with individuals with both ASD and ADHD, Chapter 5 is close to my heart. I would like to say thank you to Quinn Domiciliary (both managers and service users) for supporting me through my studies whilst I have worked as a Mental Health Support Worker since 2014.

As an advocate of Open Science, all experiments presented in this thesis have been preregistered (see osf link for all pre-registration documents: https://osf.io/6jd2p/?view_only=d90bbc8a26de4dd493513de47ab3103a) and data shared where noted throughout.

Definitions

It is important to note that when this thesis uses the term visual attention, I am adopting the approach that visual attention and consciousness are separate entities as pioneered by Lamme (2003). I will hence be referring to visual attention as just that, the sensorimotor processing of the presented environment, and not exploring a state of awareness.

Additionally, where this thesis refers to 'real-world', it is important to note the connotation of this term within this research. Here, I use the term 'real-world' to imply a

situation which was live. For example, using a mobile eye-tracker in a testing environment in a live, unscripted situation where reciprocal interaction is available. Although it can be argued that this is not a 'real' situation (due to the human being a participant), the ability to interact is comparable. Additionally, research of a similar nature adopts this term to describe such environments.

Chapter 1: Introduction to visual attention research

This chapter presents an introduction to visual attention research; beginning with the fundamentals of the human eye, then progressing to explore how gaze has been assessed in classic social attention studies. Following is an overview of ecological validity in social attention research. Visual attention within conversation is then described with an account of which cues are used and how these cues affect the timing of looks. Finally, atypical traits are discussed with their relation to eye movement in interactions. Overall, this literature review explores social gaze in more depth and acts as background literature for the experiments presented in subsequent chapters.

Physiology of visual attention

The human eye

The structure of the human eye is a complex sense organ which efficiently enables one of the most important senses. Vision is the most versatile of all the senses (Land, 2014) and efficiently enables us to understand the world in which we live.

The observable area of the human eye is mostly made up of white fibrous tissue known as the sclera. The sclera wraps around the whole eyeball and provides a protective coating. The white of the sclera contrasts with the dark pupil and coloured iris which surrounds it in the centre of the (observable) eye. The iris appears coloured due to incident light reflecting off the iris, with the colour determined by the density of melanin pigmentation. The pupil allows light and other visual information to enter the interior of the eye, while the coloured iris allows more or less light into the pupil by expanding in darkness and contracting in bright light (Rogers, 2011). The pupil has this function to stop the light sensitive cells from becoming overwhelmed. Both the iris and pupil are encased by a cover known as the cornea. The light is received by the retina (a layer of tissue situated at the back of the eye), where it is converted into neural signals. Once the light reaches the retina, the visual information captured by the eyes is transmitted to the brain. The information travels via the optic nerve to the occipital lobe (which is used to perceive information) to be processed. The final area involved in vision is the visual cortex located within the occipital lobe. Here, sensory and motor information becomes integrated with vision. This is a constant process with our eyes moving rapidly and the brain continually processing visual information, with our eyes adapting and collecting this information in parallel.

Physiology of eye movement

Saccades, smooth pursuit movements, vergence movements and vestibulo-ocular movements are categorised as the four basic types of eye movement (Purves et al., 2001). Eye movements can be voluntary or involuntary (reflexive). Saccades are fast movements which abruptly change the point at which we are fixating. Saccades can be voluntary, but reflexively occur when presented with a target (Purves et al., 2001). Smooth pursuit movements are slower tracking movements which occur when observing a moving stimulus. This type of movement allows the object to remain on the fovea. Vergence movements allow accommodation of depth and align the fovea of each eye when observing targets of different distances. Finally, vestibulo-ocular movements are used to control for head movements to keep the object in the same place on the retina. Each type of eye movement is used for different tasks to aid the gathering of visual information, with the four eye movement categories coupled with distinctive ocular movement of the muscles. The discussed eye movements are controlled by six thin muscles located between the eye sockets and eyeballs, which pull the eyes in different directions. The majority of eye movements are carried out by unconscious, reflexive processes, demonstrating the complexity and accommodating processes of the visual system. This thesis will mainly describe eye-movements in terms of fixations and saccades, in line with previous research of a similar nature.

The 'Cooperative Eye Hypothesis'

Tomasello et al. (2007) argue that humans possess, and make use of, eyes which are uniquely "cooperative". This conclusion is partly based on the finding that, compared to other great apes, humans have considerably higher contrast between their light sclera and dark iris. It has been argued that the white sclera in human eyes makes them uniquely visible to conspecifics (Kobayashi & Kohshima, 1997). Following Kobayashi and Kohshima's (1997,

2001) identification of the uniqueness of human eyes among primates and the speculation that this aids gaze following, Tomasello et al. proposed "The Cooperative Eye Hypothesis". The underlying idea of this hypothesis is that a salient eye, with a white sclera, makes it easier to infer gaze direction in others and in turn allows for shared intentionality. This, it is argued, is a fundamental cognitive trait that greatly facilitates human development and cooperation (see later in chapter for more information on gaze following). Tomasello et al. (2007) test a key element of this hypothesis, demonstrating the ability of human infants and apes to rely on eye and head movements, respectively, to follow experimenter gaze.

This is an intriguing topic when investigating the aims of this thesis, to explore how signalling cues affect attention to a speaker. As is well established, the eyes appear to be 'special', but this theory goes further to suggest we have actually evolved to facilitate our human interactions. This being that the eye physiology enables a clear signalling cue. Questions then arise as to what happens if we manipulate the presence of the eyes as a signalling cue in live social settings. This thesis explores this further with a live exploratory experiment (4) and video-based experiments (7 and 8) to assess how the availability of the eyes during conversation modulates gaze. The next section moves away from the physiology of the human eye and explores what captures visual attention.

Where or what attracts our basic visual attention?

When observing our environment, all objects are not perceived to be equally as interesting or equally capture our attention. We cannot process everything in our visual field at once, hence our attention must be directed to specific elements of our environment from one moment to the next. As humans, the visual information which we receive from the world, which is subsequently projected on to the retina, is not just a passive process. Instead, we can be active creators of how we perceive the world around us. As active participators, we can

carefully select aspects of a visual scene for further analysis and neglect other aspects of items in our visual field (Kanwisher & Wojciulik, 2000).

The way our attention is deployed to different aspects of the world around us is an efficient and sophisticated process. However, the choice of attentional focus on a moment-to-moment basis is often not a conscious decision of our visual system. This selection process is key and enables the oculomotor system to direct gaze to the optimal point in our visual field to enable efficient information collection (Kaspar et al., 2013).

The factors which derive our saccadic decisions have been widely researched in terms of inspecting static scenes (Ross & Kowler, 2013), with reference to the debate of whether the eye movements of humans are controlled by bottom-up or top-down factors. The physical properties of a visual stimulus, such as the contrast or saliency of the features would be an example of a bottom up factor of visual processing (Koch & Ullman, 1985). In comparison, top-down factors include aspects perhaps relating to the task in hand and include the perceived relevance of different locations, constraints of memory and voluntary attention (Ross & Kowler, 2013).

The neural components of visual attention with reference to top-down or bottom-up signals have been investigated by Buschman and Miller (2007) using monkeys. The authors wanted to assess the neural activity and synchrony of the frontal and parietal cortices, which previously had not been directly compared. Buschman and Miller therefore recorded the neural activities for these brain areas simultaneously during bottom-up and top-down visual search tasks. Their study provides neurological support for the belief that two separate routes are used in visual attention with top-down and bottom-up signals arising from the frontal and sensory cortex, respectively.

When inspecting the physical properties of a visual stimulus, the selection processes of visual search have been defined by a number of computational models which assess both

overt and covert attention. In the observable environment, our visual attention may be drawn to a stimulus because of its colour, shape, saliency or movement. Given this, saliency maps have been constructed which suggest there is strong evidence that salient features such as intensity, colour and orientation, are most likely to attract attention (Itti & Koch, 2001).

Within visual search research, saliency-based visual attention has often been described as a rapidly shiftable "spotlight" (Driver & Baylis, 1989;Itti & Koch, 2001). This is the understanding that attention acts as a shifting single locus which is of variable size within our visual field, whereby solely the illuminated item is processed. Alternatively, Desimone and Duncan (1995) were one of the first to take a different approach to explaining visual attention. Moving away from the view that attention functions as a mental spotlight, they instead developed a model of the underlying neural mechanisms which work to resolve a competition of visual information to process.

More recently it is understood that the visual system is a complex computational system processing information from the world around us, which struggles to be explained conclusively by one model. An attempt has been made to explain visual attention mathematically with an integration of various environmental factors. Object integrality, selective report, spatial positioning and selection criteria are a handful amongst other aspects which should be considered when explaining visual attention (Bundesen, 1990). Overall, it is understood that as humans we use a mixture of both overt and covert systems, which contain both bottom-up and top-down organizations depending upon the specific visual circumstance.

Classic social attention research

The next section gives an overview of what we understand about social attention research by highlighting classic, prominent studies from the field.

Innate social attention

From birth, we are drawn to the faces of others (Pascalis et al., 1998), implying that we are automatically programmed to be social beings. Infants as young as 10 months have even been shown to follow gaze (Corkum & Moore, 1998) and even in complex scenes, Frank, Vul and Johnson (2009) have demonstrated that eye movements tend to be directed toward faces in very young infants. This tendency has been shown to increase with age when testing infants of three, six and nine months. Farroni et al. (2002) tested new-born babies as young as 2 days old and established that from birth, infants prefer to look at faces that engage in mutual social gaze. Not only do faces capture attention (Theeuwes & Van der Stigchel, 2006), but infants have been shown to possess the ability to recognise (Kelly et al., 2005) and show preference to a mother's face (Bushnell, 2001). Furthermore, Collis and Schaffer, (1975) have reported that infant's looking behaviour can actually show synchronization patterns with their mothers. Therefore, even in the very young, this demonstrates how gaze is used for interaction tasks and is a fundamental social aspect of the visual system. Gaze therefore operates as far more than simply observation.

Classic social gaze research

Gaze has been repeatedly tested in a lab environment in an attempt to establish the effect of the eyes on behaviour. A classic study which confirmed the idea that we follow the gaze of others, even from a simple line drawing is Friesen and Kingstone's (1998) 'The eyes have it!" paper. Their study, which used a line drawing with eyes gazing to the left, right or facing forward, demonstrated how targets were faster to respond when the eyes correctly cued the target, despite being told that the gaze cueing did not predict where the target would appear. They found these results in both detection, localization and identification of the target. Friesen and Kingstone suggest this study provides evidence for covert, reflexive

attention. The findings have influenced an abundance of research which addresses whether the eyes are 'special', or whether this reflexive attention effect can be found using non-biological directional cues. Altogether the research demonstrates how following the gaze of others appears to serve a function of gaining information.

The tendency to look where others are looking has been demonstrated in simple experiments created as an expansion of the Posner (1980) cueing task. Langton and Bruce (1999) demonstrated how there is reflexive visual orienting which is triggered by the presentation of social gaze. In their experiment, a fixation cross was briefly presented in the centre of a screen, followed by a face cue. The face was either looking up, down, left or right. Following the face, a target was presented. Participants were asked to press the space bar as soon as they detected the target letter. The face shown was orienting towards one of the target locations and hence the target letter could be in a cued or an uncued location. Although participants were informed that the face orientation is not necessarily congruent with the target location, participants had significantly faster reaction times when the face acted as a cue, than when the two were incongruent. However, this finding was only seen when the cue and target interval was short (100ms). When this interval increased, the effect vanished, with Langton and Bruce (1999) concluding that this type of visual attention was a reflexive shift in line with results by Friesen and Kingstone (1998).

This has further been supported by research such as Driver et al. (1999), who demonstrated, using a gaze curing paradigm, that participants are faster at predicting targets which are gaze cued. This supports the idea that eye gaze can orient attention. However, whether eyes facilitate cueing more so than arrows of similar saliency is debated (Kuhn & Benson 2007), demonstrating how low-level saliency matters in such attentional processes.

Neurophysiological evidence supports the idea that there is a neural system which appears to be dedicated to processing the direction of other's gaze. As highlighted in

Langten, Watt and Bruce's (2000) paper, Perrett et al., (1992) used single cell recording in macaques to establish that a certain population of cells in the superior temporal sulcus region of the temporal lobe responds maximally when observing different gaze directions in others. When removing this area of the macaque cortex, it was found that primates are unable to perform gaze direction judgements, despite being able to carry out other face processing tasks (Langten et al., 2000). This has also been replicated in humans who have suffered damage to this area of the brain (Campbell et al., 1990).

Overall, it seems the eyes of others influence our behaviour in classic social gaze research. The next section explores why the face as a whole attracts visual attention.

Function of looking at the face

The reasons as to why the face receives attention have been articulated in early work by Kendon in 1967. When engaging in conversation with a partner, Kendon explains this form of gaze serves at least 4 functions. These include providing visual feedback, to regulate conversation flow, to communicate emotions and restrict visual input to help improve concentration (Vertegaal et al., 2001). Although there appears to be some form of learned behaviour from others about where and when to look during a social interaction, from previous developmental research, the instant attraction appears to be an innate behaviour.

Interestingly, atypical populations, such as individuals with Autism Spectrum Disorder (ASD), demonstrate discordant eye movement patterns when comparing gaze in a typical population. Chawarska and Shic (2009) used an eye tracking procedure to examine the visual scanning and recognition of faces in 2 and 4-year-old children of typically developing children and children with ASD. Their findings demonstrated that with age, the ASD children increasingly looked away from the faces, with this linked to increasingly impaired facial recognition. Looking at faces of other individuals therefore appears to be

crucial for recognition and effective social interaction. This research has influenced Experiment 7, where social attention to faces is explored in high and low traits of ASD. See section 'Atypical traits' for further research on gaze behaviour in ASD traits.

Furthermore, in order to understand others intentions and emotional state we look to the face for this information with Ekman, Friesen and Ellworth (2013), describing the role of the face in social life. They suggest our facial muscles allow more than a thousand different facial appearances, and these different displays allow us to send messages about our feelings on a moment-to-moment basis. This therefore makes this area of the human body a rich information source for others who we engage with.

The distinct processing of eyes

Further support, that the face is indeed a distinct feature when analysing human gaze, is present in a review by Langten et al. in 2000. The review discusses models of human face processing in communicating information and in particular the importance of eye gaze. The authors explain that there is a neural mechanism which has evolved that is devoted to gaze processing.

When analysing the way faces are processed, Keil (2009) investigated the key internal features which are crucial to identifying faces. The results showed that for recognition of faces, humans rely on a narrow band of spatial frequencies. The research demonstrated that the eyes and mouth give the most reliable signals for facial recognition. The author infers that the brain has built-in knowledge of this ability. This further supports the belief that eyes are 'special' and maintain our focus in social interactions.

As the eyes are recognised as important to attention in social interaction, Peterson and Eckstein (2014) investigated the reasons for this and considered whether this behaviour is caused by social importance or whether it has a functional importance. They proposed there

are optimal fixation points which are specific locations that we attend to. However, these fixation points relate to the context of the situation and vary according to the task given. For example, the fixation points differ according to whether we are attempting to identify the targets' identity, gender or emotional state, demonstrating how the specific environmental circumstance needs to be considered. Overall, they found observers tend to fixate just below the eyes with these findings replicated in studies including those simulating a visual deficit (Tsank & Eckstein, 2015).

The eyes of others appear to be a key feature that captures visual attention in a social context. During human interaction, we focus on the eyes in an attempt to understand others' intentions, beliefs, and emotional states with the ability to follow and direct the attention of others. Additionally, the belief that the eyes are processed distinctly to other aspects has been supported with accompanying neurophysiological research. The distinct processing of the eyes generated ideas of manipulating the availability of the eyes for experiments (4,7, 8 and 9) presented within this thesis.

'Ecological validity' from lab to real-world

So far, this introductory literature review has discussed how visual attention is directed, what gains our attention and why looking at the face is distinct. These studies are classic lab studies which have their importance to understand lower-level features. However, when attempting to understand the complexities of gaze in social situations, we must assess the research which expands on classic lab studies and the ecological validity considerations. Social attention research has to consider the degree to which stimuli can be controlled, when balanced against the amount the experimental design resembles a real-life situation. This next section highlights the problems with ensuring ecological validity in terms of social attention

research and further features additional factors which should be considered when attempting to generalise results to everyday life.

Coining the term 'ecological validity'

When we refer to the term 'ecological validity' it can often be used as a piece of throw-away vocabulary to criticise a paper. The extent to which a piece of research is termed ecologically valid depends on the user's definition of the term.

Egon Brunswick was the first to use this term in 1947 (Brunswick, 1947). His definition was based in the field of Perception to describe how a proximal (e.g., retinal) cue relates to a distal (e.g., object) variable. In his example this was to understand how representative a design is. In a more recent paper, Hoc (2000) explores the ecological validity of research in cognition and defines the term as: *"a possibility to generalise the conclusions obtained by the study of an artificial situation to a class of natural situations"*. Although for many of us, this would be a fair description of how we use the phrase, the term, however, does appear to have some problems in its accepted definition (some of which are discussed next).

A problem worth probing when trying to define the term 'ecological validity' is deciding which aspect of the research we are basing this judgement on. For example, according to Schmuckler (2001), this term could relate to the type of stimuli, the response to the task and/or the nature of the research context; giving the evaluation of ecological validity at least three dimensions.

In current research by Holleman et al. (2020), they explore how the social attention literature uses the term with a range of ambiguous definitions. They describe this as *"The Big Umbrella of Ecological Validity"*. One influential definition Holleman et al. (2020) cite, is that by Brofenbrenner in 1977. Here, Brofenbrenner describes ecological validity as: "*The*

extent to which the environment experienced by the subjects in a scientific investigation has the properties it is supposed or assumed to have by the investigator". Holleman et al. (2020) explain this, moving forward that *"…theoretical considerations should guide one's methodological decisions on what type of research context is most appropriate given one's focus of inquiry"*. In other words, rather than striving to meet the requirements for greater ecological validity, we should first look to our research question and ask ourselves, as researchers, which method is most suitable.

Alan Kingstone's advocation of 'Cognitive Ethology' (Kingstone, Smilek, and Eastwood, 2008), is of particular importance for social attention research specifically. This refers to the belief that research should first begin with a real observation of behaviour that is seen in the real-world. This behaviour should then be taken into the lab to explore further, and not the other way around. Despite this, in Kingstone's (2009) paper, he highlights how often studies involving eye movements fail to adopt Cognitive Ethology. Kingstone explains the importance of this as often lab-based studies are so simple and controlled that they do not take into account the situational complexity of social attention.

Hoc (2000) considers that there is a fragile trade-off between the research costs and relevance when trying to obtain ecological validity in methodological design. The authors state that often field research with costly methods could have been conducted in laboratory settings with less cost. In Experiment 5 (Chapter 4), I demonstrate how third-party visual attention to videos of groups is very similar to live observation. This is a comprehensive example of how a lower cost method (third-party lab testing) can be equivalent to live 'field' research, which is often less efficient. In terms of social attention research, it is imperative that this trade-off is carefully considered.

Factors to consider in social attention research

When contemplating the extent to which social attention studies are ecologically valid, we may focus on how often and even *if* the findings from these studies can be seen in a real-world situation. A key component of 'real world' research is the type of stimulus used.

Risko et al's (2012) highlight that in recent years the methodology of social neuroscience research has been criticised due to the absence of real social encounters. In order to map the social brain, the stimuli used in research of this context are simple, static representations of socially relevant stimuli rather than actual live interactions. Hence the assumptions about the brain which are gained via these methods are argued as dangerous and sometimes misleading (Schilbach, 2010; Kingstone et al., 2008). Risko et al. (2012) instead offer a comparison of different types of social stimuli which range in their approximation to real life. They assess a number of different studies, which use a range of stimuli along a continuum of this scale ranging from schematic faces through to real social interactions. They argue that it is possible to assess the similarities and differences of how we attend to the social stimuli through direct comparison of various stimuli (e.g., how we attend to a schematic face versus looking at a real face).

The first type of research highlighted in the review is the classic attention orienting example by Posner (1980), which was adapted to explore gaze cueing by swapping the arrows for eyes (Friesen & Kingstone, 1998), as previously described. This research is renowned and has been particularly influential in this field of psychology, primarily due to the robust gaze cueing finding. Risko et al. (2012) demonstrate how gaze cueing is supported by neuroscientific research. However, when comparing the schematic faces used by Friesen and Kingstone with real faces, Sagiv and Bentin (2001), demonstrate the brain responds differently to the two pictures. Risko et al. (2012) argue that gaze processing can therefore be overtly different even with a small simple difference of stimuli. This highlights a key consideration for social attention generalisations.

Importance of using moving images

One should also consider the type of images used. When attempting to understand visual attention in the dynamic world around us, it is important to progress from static stimuli. Ross and Kowler (2013) argue that videos can be a useful tool to identify the factors which influence our saccadic decisions in natural situations. Using video clips with eye tracking technology enables a dynamic scene to be examined, advancing from static classic social attention research. Ross and Kowler (2013) also articulate that the use of videos allows a comparison in the performance of the study of eye movements when the content is identical across observers.

Although there are benefits to using video clips to examine eye movements, Dorr et al. (2010) demonstrate that the type of clip that is used is an important factor which needs to be considered when designing social attention stimuli. They suggest when using Hollywood action style movie trailers, which are designed specifically to direct attention to a specific location, there is more coherence in eye movement when comparing variability across a large data set. The authors also examined the difference between static images and dynamic continuous videos with static video observations driven more by stimulus onset effects. Dorr et al. (2010) conclude that these types of clips which are often used as eye tracking stimuli in laboratory-based experiments are not representative of natural viewing behaviour. Their results encourage further research on viewing dynamic natural scenes with reference to the different observation patterns observed in varying stimuli. Therefore, future research should address and consider the purpose of the video stimuli selected. This consideration is adopted within experiments in this thesis.

Social presence

Gaze in a real-world situation can be more complex than third-party viewing in a lab, with gaze location in a live situation altered by a number of factors, including social

presence. The extent to which the third-party viewing and live situations are comparable has been researched in studies including an important piece of research by Laidlaw et al. in 2011. The study involved measuring participants' looking behaviour as they were sitting in a waiting room with another confederate or with a video recording of that confederate. The study explored to what extent the participant looked to the confederate in both situations. It was established that looks to a live confederate were less than when the confederate was an image playing via videotape. Therefore, the physical presence of this confederate affected the participants' gaze location, something which should be considered when attempting to generalise results from lab to life. A further interesting point when considering divergence in behaviour from these two situations is 'civil inattention'. Goffman (1966), used this phrase to describe how strangers have an unspoken mutual agreement to not engage in an interaction. Zuckerman, Miserandino and Bernieri (2011), have demonstrated how civil inattention exists in elevators, where we tend to avoid eye contact with our fellow passengers. It therefore seems that when in close proximity, under certain live situations, we tend to refrain from looking to others. This is further emphasised in a recent study by Mansour and Kuhn (2019). Using a novel and clever set-up, their research demonstrated how visual attention is deployed to different areas in live versus third-party situations. The authors created an experiment whereby half of participants were told they were engaging with a 'real' Skype call and half were told the stimulus was a pre-recorded video. Key differences were found in the position of gaze, with those engaging in a 'real' interactive call looking less to the eyes of their interlocutor. Mansour and Kuhn conclude that social attention processes are significantly influenced by the nature of the social interaction.

Gobel, Kim and Richardson (2015), demonstrated how manipulating the beliefs about the social context of a situation modulates gaze. Participants were asked to look at targets of perceived high and low ranking. Counterbalanced between participants were two conditions:

one-way viewing and two-way viewing. In the two-way viewing condition, participants were told the participants themselves were being recorded and their videos would be watched at a later stage. Those in the one-way condition were told the recording was for measurement verification purposes, seen by the experimenter, and only if required. Findings demonstrated that eye movements did change depending on the participants beliefs. When participants believed the targets would later be looking at them, participants looked less to higher ranked targets' eyes. In other words, looking behaviour varied according to whether the participants gaze was being used to perceive or to signal (belief that they too are being viewed). Hence validating that gaze is modulated by social presence has a dual function.

With a similar manipulation, Gregory et al. (2015), demonstrated the impact of social context on social attention. Their study involved three conditions of varying social settings. In two groups, participants believed they were watching a live webcam of other participants, in one of these groups, they were informed they would later complete a task with these participants. In a third group, participants simply viewed the scene freely. The researchers found that in the two groups where participants believed they were watching a live recording, there were less looks to the heads of the people in the scene and less gaze following.

Furthermore, in a recent study by Holleman et al. (2020), who used similar 'live' versus pre-recorded conditions, they too found that participants in the 'live' situation gazed less to the eyes of the confederate than in the pre-recorded setting. The authors suggest that the social context therefore modulates gaze.

These highlighted key differences in implied live social presence versus a non-interactive context may help to explain Laidlaw et al's (2011) findings and other described differences, in that social norms play a large role in our gaze behaviour when in the presence of others.

Gaze as an information signal

During a social interaction, it is important to note that an additional complexity is that, in a face-to-face interaction, both you and your partner are able to follow the gaze of each other and in turn, signal. Hence gaze not only allows us to gain information from the world around us, but also, we are able to use others gaze to inform us.

Perspective taking and gaze following

The way in which we follow the gaze of others, links to Theory of Mind (ToM) and perspective taking. When observing gaze of others, we are aware of what they can or cannot see and perceive. The ability for adults to think about others mental states is a dedicated automatic process (Santiesteban et al., 2014) which recruits a fast and efficient processing system (Nielsen et al., 2015). This is demonstrated during the dot perspective task. When participants are presented with an avatar who can see incongruent information to ourselves, perspective-taking judgements reaction times are affected (Nielsen et al., 2015), which is also verbalised as 'altercentric intrusion effects'. Nielsen et al's (2015) results indicate a significantly stronger effect when the experiment involves a social context (e.g., involving a social agent compared to a colour block). Whether this is a spontaneous effect or not has been debated in work by Cole et al. (2016). This study expands upon the dot perspective task but adds the additional manipulation of blocking the persons observed vision. Findings were that there continued to be an effect even when vision is obscured. Although their results refute previous automatic perspective taking results, the authors do not argue that this task does not involve any social processing mechanisms. This line of research demonstrates how the influence of implied social presence affects the way in which we distribute our attention, which is crucial to understanding real interactions.

The ability to infer the beliefs of others and the effect of this on gaze has been researched in a study by Crosby, Monin and Richardson (2008) who examined eye movements during potentially offensive behaviour. Their study used an innovative method where participants were presented with a pre-recorded video of 4 discussants with two conditions in which all discussants could or could not hear each other. The video featured a white target vocalizing a viewpoint which could be perceived as potentially offensive to the black target present in the video. As predicted, the black individual was looked at more than the other discussants in the headphones on condition. Crosby et al. (2008) included the headphones off and on condition in an attempt to examine whether this behaviour occurred due to association or social referencing. As the two conditions showed significantly different results, it is assumed that the looking behaviour occurred to assess the black individual's reaction when they can hear the comment.

The way our gaze is affected by what we infer from other's perceptions, has been explored by Boucher et al. (2012) using human to human and human to robot interactions. The study involved a cooperation task with two agents who must work together to achieve a shared goal. The robot allowed the authors to manipulate which signalling cues were present. Gaze was manipulated in three conditions full gaze (head and eyes present), eyes hidden (with sunglasses) and with a fixed head. Their results expand upon the idea that in human-to-human interactions, an agent's vision of partners gaze can improve performance during a cooperative task. The results not only allow a more human-like social ability in robots, but the authors argue that findings also demonstrate the pertinence of these physical cues to facilitate cooperation.

Theory of mind is explored further within Experiments 1, 2 and 3.

Joint attention

Joint engagement is typically defined by joint processing of information and sharing attention with another individual (Mundy & Newell, 2007). In a developmental population, this can be seen by assessing whether the infant shows a pattern of alternation of gaze between an object and an adult's gaze. If an infant is aware of the others focus to the object by observing their gaze and in return they respond by also engaging with the object, this can be described as successful joint attention (Carpenter, Nagell & Tomasello, 1988). In 1995, Tomasello proposed that an infant's understanding of others and the emergence of joint attention is the fundamental principle to later understand ToM. Furthermore, it is well understood that developing the ability to jointly attend to an object is considered to be critical to early cognitive and social development (Mundy, 1998). Possessing the ability to participate in joint attention and follow gaze with another individual has therefore been proven to be a fundamental developmental function and somewhat a 'social intelligence' which enables effective communication later in adult life.

Utilizing gaze

Not only has following gaze been proven to be key to development, but research has shown that the eyes are an essential tool to gain information in a social setting throughout life. Gaze cue utilization has been investigated in a real-world situation by Macdonald and Tatler (2013). Their study involved an instructor explaining simple tasks with building blocks to the participants in a face-to-face live situation. The instructions were either unambiguous or ambiguous and the instructions given either included or excluded gaze cues. These cues were implanted by the instructor either directing his gaze at the correct building block or reading the instructions whilst looking at the paper they were being read from. Findings suggest that if gaze cues were present, the participants on average had a more accurate performance, with more fixations towards the instructor when unambiguous instructions were

given. Their conclusions include that during an interaction such as this, we seek gaze cues of others as a means to provide unavailable information. In other words, in social situations where we are unsure of the task at hand, we tend to inspect the gaze of other present individuals, to reveal the correct behaviours. This acts as a form of informative social influence.

It is therefore apparent that we can view the gaze of others as a means to understand the social setting in our environment. This emphasises the importance of gaze and supports, with an abundance of evidence, the belief that there are multiple functions of gaze that expand beyond simple looking behaviours.

Eye movement in conversation

In terms of this thesis, it is important to think about how this may interplay with an interaction which involves a spoken conversation. The aspects which affect gaze have been discussed, but during conversation, when exploring eye movement and visual attention, we see a combination of all of these aspects, something which makes this type of research more complex.

During a conversation, there are many parts of a scene which may attract or influence where a person's visual attention is directed. For example, whether it is a live conversation or third-party video observation. In both cases, we have moved from a static image to a dynamic scene. The people within the conversation will move, react, express emotions, and take turns to speak. They will use their faces, hands and body and voice to direct attention. Using people as stimuli increases the unpredictability and extraneous variables which makes conditions difficult to measure objectively, opposed to schematic images of faces. In addition, if we are examining live face to face interactions, this also means the person is dynamically engaging rather than passively viewing, adding to the complexity of

conversation research. A key aspect of conversation to consider is that it combines audio and visual modalities. This means attention is often shared to other areas of the person such as the mouth, rather than the eyes persistently dominating which is described in classic social attention studies.

During speech, it would be plausible to assume your interlocutor would look to your mouth, as would you theirs. However, if you imagine yourself having a conversation with a friend and they persistently focus on your mouth, this will violate our social norms. In fact, it may even be seen as an advancement with an abundance of 'pop' psychology determining that looking to a person's mouth equates to an attraction to that person. Equally, it may be extremely off-putting for the person speaking, making them feel self-conscious. The reality is, that during a normal conversation, we tend to move our eyes rapidly between the eyes and mouth and nose of others, focussing on the eyes as centre of the face. Vo et al. (2012) demonstrated how, when a face is moving quickly, we tend to use the centre of the face as a spatial anchor. This effect, however, can be manipulated. An abundance of research has demonstrated how there are increased looks to the mouth area if the conditions make it harder to understand the person (Vatikiotis-Bateson et al., 1998; Buchan, Paré, & Munhall, 2008). This makes sense, if you think about a time that you have been in a busy room at a conference where you cannot hear the person well, you may look to their mouth to aid your understanding through lip reading.

Eye contact

In conversation, eye contact is important for successful communication. In a review, Kleinke (1986) demonstrated how eye contact is important for many aspects of interaction, including but not limited to: regulating interaction, providing information, expressing intimacy and facilitating goals. Additionally, eye contact has been shown to aid

understanding (Clark & Brennan, 1991) and enhance perceived speaker credibility and honesty (Beebe, 1974).

It is important to study eye contact when trying to better understand behaviour in a live interaction. This is due to the dual functionality of the eyes. As previously noted in this thesis, the eyes not only take in information but also signal information. Therefore, to fully understand the function of the eyes, it is important to engage in methodologies which allow for this interaction to occur, rather than static images. Examples may be using video which allow for (implied) eye contact or in live face to face interactions.

In terms of eye contact in live interactions, studies using brain imaging techniques have highlighted the differences observed in brain activity when engaging in an interaction with different eye contact conditions. For example, Hoehl et al. (2014), explored oscillatory brain activity in 9-month-old infants when looking at an object with an adult in two conditions: with and without eye contact. When eye contact was present, the infants responded with a desynchronization of alpha activity. Myllyneva, Ranta and Hietanen (2015), explored brain responses in direct gaze for individuals with social anxiety disorder. Results indicated that there were greater responses when viewing faces with direct eye contact versus averted and that interestingly, the responses to the eye contact were only enhanced if the participants believed they could be seen. This arguably adds a ToM element to the research, in that increased brain responses were only present when participants are aware another person is aware of them. It is therefore apparent that eye-contact has not only a psychological but also a neurological effect.

Freeth, Foulsham and Kingstone (2013), explored the effect of direct and averted eye contact on gaze patterns. Their study's testing session allowed for a natural conversation flow with the use of a mobile eye-tracker to record eye movement. The experimenter manipulated their eye contact in two conditions. In one condition, the experimenter engaged in

conversation by making direct eye contact with the participant. In a second condition, the experimenter averted their gaze to look down at their notes, and these two conditions were counterbalanced to ensure there were no order effects. The researchers reported that the experimenter's gaze did influence participant viewing behaviour. In particular, in a live interaction, participants looked more to the experimenter's face during eye contact. Additionally, increased autistic traits were associated with less looking at the experimenter but for video interactions only.

Eye contact has also been studied in video settings. A study by Doherty-Sneddon et al. (1997), explored the effect of eye contact in conversational turns, with 3 conditions: video with and without eye contact and audio alone. Dialogues in the video condition which enabled eye contact resulted in participants taking significantly more turns of talk and uttered more spoken words. The authors concluded that if a video enables eye contact, then it can increase the amount of verbal interaction compared to when eye contact is not possible. The authors suggest that gaze is vital for providing interpersonal feedback.

Arguably this is not 'true' eye contact as it may not be direct and is not presented in a face-to-face situation with live social presence. This is particularly interesting in terms of disinhibition which is often experienced in an online environment. Termed 'the online disinhibition effect' (Suler, 2004), this describes the behaviours of individuals who lack restraint when engaging online rather than face to face. Suler (2004) has suggested that this may be due to the absence of eye-contact during these online interactions.

In a review by Senju and Johnson (2009), they explore models for how eye contact is processed within the brain (which can explain some differences in ASD, see subsection on atypical traits). The affective arousal model proposes that eye contact activates brain arousal systems directly, which in turn elicits an emotional response. A second model highlighted, is the communicative intention detector model. This model links to theory-of-mind, where it is

suggested that eye contact signals intentions to communicate. A further model is the fast-track modulator model. This described eye contact processing as "quick and dirty" (LeDoux, 1996) in that the route is fast, with low spatial frequency, operating subcortically.

Estimating eye-contact

Recent work by Müller, Sood and Bulling (2020) has provided us with a method to predict when eye contact will be held and when gaze will be averted. Using multiple factors of multimodal social signals (such as speaker diarization, head pose and facial expressions), their research fills a gap of being able to forecast eye contact during everyday conversations. The authors suggest their work provides evidence for the interplay of eye contact and other non-verbal signals and takes an important step towards methods for anticipatory interactions. Using YouTube videos to assess the reliability of their approach, they suggest their method allows for a more accurate prediction than comparable baselines.

In a recent study which explored how often mutual eye contact is established, Rogers et al. (2018), found the amount of mutual eye contact during a four-minute conversation with an acquaintance equated to only roughly 0-45% of the time, with there being large discrepancies between participants. Equally, the time frame of these durations was extremely short, with mutual eye contact exhibiting as brief instances. Despite this, after taking part, the experimenter asked the participants to estimate the amount of eye contact they made. Surprisingly, participants thought they made eye contact more often with estimates of around 70% of the time.

In line with a lay perspective, we value locking eye contact as desirable and beneficial to our successful interactions. The above research suggests perhaps eye contact doesn't occur as much as one would estimate. Instead, perhaps this overzealous approximation comes from the fact we are aware we are monitoring eye movements, by continuing to flick back to our

interlocutor, rather than maintaining eye contact. Experiment 4 offers an explorative account of the effect of eye contact during conversation.

Turn taking signalling

Using a live setting, Ho, Foulsham and Kingstone (2015), explored the use of dynamic gaze as a signal during face to face interactions. Their study expanded upon the belief that gaze can be used to control turn-taking behaviour by analysing the temporally sensitive characteristics involved. The study aimed to provide an insight into how dyads partake in turn-taking behaviour on a moment-to-moment basis in a more natural interaction. Their findings validated previous evidence which suggests the use of gaze as a signalling mechanism. Furthermore, the research demonstrated the spatiotemporal patterns of gaze patterns during dyad conversation. The authors report how speakers ended their turn with direct gaze at the listener and when the listener begins to speak, they do so with averted gaze. A particular advantage of this study is the use of a natural setting, with two interacting participants, taking the participant away from a computer screen and providing a glimpse into social attention in the real world. The temporal characteristics of this study are discussed later within this chapter.

Hessels et al. (2019). attempt to quantify the complexities in gaze allocation during live communication. Building on Ho, Foulsham and Kingstone's (2015) findings, they explored joint gaze with a view to understand how task structure modulates gaze behaviours. Using a sophisticated set up which allows for dual eye tracking with direct eye contact between dyads, they found that gaze was modulated by the task at hand (speaking versus listening), whilst also confirming the turn-taking gaze 'dance' that interlocutors display.

Hautala et al. (2016) explored how different elements of conversation have an effect on the way third-party observers view that conversation. Their study explored the effect of

the conversation content on observer gaze behaviour. Participants viewed two actors engaging in matter of fact or personal conversation. Their hypothesis was that in addition to turn-taking behaviour, the semantic content may also modulate attention location. They predicted that observers gaze would synchronise with the conversation structure, in that their main focus would be with the current speaker, but also that the interpersonal speaker would cause increased fixations with the second speaker or a faster shift of attention toward that speaker. They found that during the first spoken line, as expected, attention shifted to the first speaker as they began to articulate. Interestingly if the person made a negative personal comment, the attention shifted to the second speaker. The authors suggest participants may be looking to the (potentially aggressive) reaction to the provocation as a potential reason for this gaze shift. Furthermore, in personal versus matter-of-fact conversations, the second speaker seemed to captivate the observer's attention. Results indicated that the semantic content of the conversation does dynamically modulate the spatiotemporal gaze patterns of observers. This links to top-down factors of visual attention as previously discussed.

Group size

A challenge of social research is the ability to analyse looking behaviour in larger groups. The majority of past research includes dyads, with only a handful of experiments exploring gaze behaviour in larger groups. Speculations as to why groups have not been investigated as often as dyad pairs may include the complexities of obtaining and analysing data when there is more than one other person to respond to. When an additional person is added into the study, not only does the way they behave need to be analysed, but this also creates additional interactions and with that it equates to additional variables which can be harder to control and analyse.

For example, in a dyad interaction, there are only two choices, you are either looking at the person, or you are not. A common eye-movement pattern of a dyad pair generally involves the listener looking directly at the speaker (perhaps to show they are engaging with the conversation). Conversely, the speaker begins their turn of talk by averting their eyes away from the listener. The speaker may glance back to the listener to gage their response, but ultimately spends most of their utterance visually locating elsewhere. Interestingly, when the speaker finishes their turn, they tend to gaze back to the listener, almost like a subtle cue to signal they have finished and are ready for a response. Analogously, when the listener becomes the speaker, there is a strong tendency to avert their gaze upon beginning their turn (e.g. Ho, Foulsham & Kinsgtone, 2015).

Arguably, a dyad pair enables a researcher to assess with ease when a person is looking at their partner or not. However, adding additional members to the conversation, and hence moving the interaction partakers from a pair to a group, means the addition of further complications.

For example, now rather than having just one key area of interest for visual attention (the other half of the dyad pair), there are now two or more. In a triad, you, as an observer must decide whether to look at someone, but also *who* you should look at. For instance, do you focus the speaker, or should you look to the third person to gage their reaction (social referencing)? This makes the dynamics of a group conversation more complex.

As discussed, in a dyad pair interaction, it has been established that we engage in a sophisticated pattern of eye movements to signal our turns of talk, but how does this change in a triad? We are now interacting with an additional person, so are you sharing an 'eye movement dance' with person two *and* three? And are person two and three also sharing a 'dance' amongst themselves?

Turn-taking speech also means a participant within a group has a more complicated decision-making process of when to speak and when to listen, resulting in more occasions of people talking over each other. Not only this, when we move from a dyad pair to a group, there tends to be increased social interactions which often results in a dominance pecking order. For example, personality traits and mood will play a part in who controls the conversation and there is an increased chance of one group member becoming less involved. For this reason, it is clear why researchers predominantly assess their research questions with a dyad pair and why our knowledge of visual attention deployment in group settings is shallow. We cannot assume that the visual behaviours are equivalent when adding additional members to the conversation, due to the dynamic and complex nature of a group interaction.

There is some limited research to date which has demonstrated that the group size may affect how visual attention is deployed, which will be discussed next.

Vertegaal et al. (2001), conducted one of the initial and limited experiments into eye movements in groups (where the targets in the stimuli include more individuals than one dyadic pair). Their early study involved seven four-person groups, sat at a table discussing current-affairs, with subjects' eye movements recorded with a desk-mounted eye-tracker. Analysis involved combining the speech data for conversational analysis and gaze data which meant percentage of time spent gazing at each partner while speaking or listening could be calculated. Their results demonstrated that participants looked more at the person they were speaking and listening to. Hence, the authors describe gaze behaviour as a predictor of the user's conversational attention. Their research also confirmed the differences of gaze behaviour in larger groups. The study showed that total amount of gaze increases when addressing individuals when within a triad opposed to a dyad. Their paper suggests three reasons why speakers look more when addressing larger groups: visual feedback (when speaking to a triad versus a dyad, there is less time to collect feedback on each individual,

hence more gaze), communication of conversational attention (used to maintain a signal of who you are speaking to), and regulation of arousal (in that when addressing triads, a speaker would need more gaze to maintain sufficient eye contact).

Vertegaal et al's (2001) final hypothesis took into account the amount of time spent gazing at individuals listened to, in comparison to time spent gazing at an addressed individual. Findings were that subjects gaze on average 1.6 times more when listening than while speaking, which is in line with other dyadic research. Vertegaal et al. (2001) conclude by stating that the gaze of listeners is somewhat more predictive than gaze of speakers when analysing conversational attention. Their results can be extended to use eye gaze for conversational systems that need to establish who is speaking or listening, by using gaze as a valuable source of input.

More recently, Maran et al. (2020) has explored the differences in gaze behaviour from a dyad pair to a group of five. Their study has strengths in the methodology in that they used a 'real' social interaction, allowing for the effect of gaze in two functions: receiving and signalling. The authors were also interested in how speaking and silence affected gaze patterns, hence they included two conditions. First, participants were asked to sit in silence and second, to engage in conversation. Their research interests and hypotheses were threefold. First, they hypothesized that gaze would be directed more to participants when engaging than when in silence. Secondly, that gaze would be directed at others more so when speaking than when listening (in line with work by Ho, Foulsham & Kingstone, 2015). Thirdly, that when engaging in a group there would be less gaze towards those participating than when in a dyad pair. All of these hypotheses were proven.

Specifically, in terms of the effect of gaze on group size, findings show that, when engaging in conversation, there was an increased display of social attention when participants were in a dyad pair than a group of five (contradictory to Vertegaal et al., (2001)). In other

words, group size did modulate gaze behaviours. This may link to a term described as 'social loafing', which describes the behaviour of individuals who exert less effort on a task when they are in a group setting (Maran et al., 2020). For example, in a dyad pair, it is important to show we are listening and engaged with the conversation, and we do so by signalling with our eyes and making eye contact. However, in a group this normative pressure is reduced, almost as if we share the workload to show our interest in the conversation.

Gaze in group conversations has also been explored using video conferencing. Vertegaal and Ding (2002) used a set-up where two actors engaged with a participant via video chat, similar to what we now know as Zoom. The authors wanted to investigate whether increased gaze to a participant allows for a clearer understanding of whether the participant is listening or speaking. Hence, whether this signal of 'I am listening' equated to the participants feeling more comfortable and more speaking from the participant. To investigate this, participants took part in a conversation in two conditions. In this first condition, participants experienced gaze which is synchronized with conversation attention (what we would believe to be 'normal'). In the second condition, the gaze by the actors was random. Their results indicated that speech was modulated by gaze, with a 22% increase in speaking in the first condition.

However, they also report that these results were due to the amount of gaze (rather than the synchronization of gaze). Correlations between amount of gaze and amount of speech were .62. Hence, the authors suggest their results do not indicate that people are more likely to speak when gaze behaviour of their interlocutors is synchronized. Instead, an increase in gaze from another person resulted in an increase in speech. This work has implications for the current climate with the global pandemic Covid-19, where the majority of our interactions now take place online.

Overall, despite there being limited evidence directly comparing dyad and larger group differences in gaze behaviour, it seems that group size may have an effect. It appears there are differences in visual attention when an individual forms part of a larger group membership. As dyad pairs are most commonly explored, future research should be cautious when making generalisations from such settings to group behaviours and whether previous findings using dyad pairs extends to larger group conversation. Understandably, the field must begin with simple conversation (for example with just two people) to understand determinants of gaze behaviour. However, we are at the stage where research can begin to expand upon these findings. Aspects of this thesis research extend this and focus on what captures our attention with multiple individuals in a group setting (for example, Experiments 4-8). It will be interesting to analyse gaze in this larger group context in an attempt to confirm if results are in line with previous findings of dyad pairs. For example, whether larger groups also engage in an eye movement 'dance' to signal to others when we have finished speaking and signal turn taking (Ho, Foulsham & Kingstone, 2015). If we extend this to a larger group, where do the eyes look to signal this to the other individuals? Furthermore, when third-party observers view interactions of multiple individuals, where do they attend?

Audio and visual cues

The next section explores audio and visual cues during conversation.

Using audio cues to guide visual attention

Imagine you are in the cinema with friends. The sound of the film is playing extremely loudly, someone in the row in front of you is rustling their sweet packet and your friend leans into talk to you. Without realising, you are able to selectively attend to the sound

of your friend's voice. When engaging in such conversation it is true that your friends eye movements, head movements and gestures are combined to assist your understanding of what they are trying to tell you. However, you are still able to selectively attend to the auditory component of their voice, whilst ignoring other background noise. This might be apparent when you then turn back to watch the film, realising you have missed a vital few seconds!

The ability to selectively attend in such situation, often termed 'the cocktail party affect', was first described by Cherry in 1953. The author demonstrated various factors such as the sex of the speaker, voice intensity and speaker location which affect our selection of attention. Relating to this, in more early research, Broadbent (1956), famously put forward his filter theory which suggests a way to process and prevent overload of multiple auditory inputs simultaneously. Broadbent suggested that if two messages are presented in parallel, they are held in a sensory buffer which holds the information for a short period of time. One input is then able to progress through the filter while the other is briefly stored in the buffer for later processing.

It is apparent that we are able to selectively attend to different audio inputs within our environments. However, how does this affect who or what we visually attend to during natural social interactions and how is sound used as a cue to guide our gaze?

The effect of audio on gaze viewing has been examined in speech intelligibility tests with the belief that the combination of audio and visual provides the best understanding of conversation. Sumby and Pollack in 1954 were one of the first to establish that seeing a speaker's face enhances speech in noisy acoustic environments. When thinking back to our cinema scenario, we can imagine this to be true.

A number of recent studies have investigated gaze using pre-recorded clips of turn-taking in conversation (Foulsham, et al., 2010; Tice & Henetz, 2011; Foulsham & Sanderson, 2013). These studies have begun to explore how the behaviour of the actors within the clips

affects the gaze of third-party observers and the effect of audio on visual attention can therefore be examined using pre-recorded video clips, which are shown to participants at a later stage. A benefit of using videos and third-party viewing, is the ability to manipulate the audio cues and track visual attention. For example, research by Quigley et al. (2008), determined that eye-movement is biased toward a part of the visual image which directly corresponds to the source of the sound, demonstrating an audio and visual association. Furthermore, research by Vatikiotis-Bateson et al. in 1998, assessed eye movements during audio-visual presentations of monologues. Findings were that when participants were given a speech task, the number of transitions between areas of the face decreased in the presence of noise. This suggests that the visual speech information is important during eye movement analysis when observing other individuals speak. Furthermore, the results suggest that the importance of visual speech information may actually affect how the information is gathered by the observer.

Similarly, in Vo et al's (2012) study, the authors investigated the effect of audio on interviews with pedestrians. Their study involved participants watching two versions of the same video clip, with the second clip containing only visual information with the sound removed. Their findings were that removing the audio speech decreased fixations to the face with the suggestions that, in this instance, there was not a general bias to look at the eyes. Instead they suggest that attention is directed to locations perceived to provide useful information on a moment-by moment basis. Vo et al's (2012) findings support the view that gaze is used in an efficient process to gain information from others.

Foulsham and Sanderson (2013) further explored the effect of removing sound when observing targets in a group setting. Their study involved participants watching pre-recorded clips of four targets engaging in a natural discussion. The study included 28 participants who were asked to view twenty 30 second video clips while their eyes were tracked. The stimuli

which participants watched comprised of four individuals sat behind a desk engaged in a non-scripted group discussion. The clips were prepared from a 15-minute video recording where the target individuals were asked to discuss and decide on an answer to multiple survival-related questions. The authors included the manipulation of removing the sound in order to investigate whether auditory information would affect when speakers were fixated, how fixations between different observers were synchronized and the number of fixations on the eyes and mouth. Differences between regions of the face were investigated in the expectation that removing the sound might increase looks to the mouth if participants were trying to decode speech from lip movements alone. The results demonstrated that removing sound led to decreased attention towards the current speaker and instead more time was spent looking at the other non-speaking targets. There were increases in looks to the mouth upon removing the sound, however, the eyes continued to attract most of the fixations. Despite these changes in gaze behaviour, the participants still appeared to follow the conversation without audio. Hence, as participant's gaze patterns continued to follow the turn-taking conversation without any auditory information, it appears the participants were using a visual cue to guide their gaze to the speakers, and that this strategy may have helped follow the depicted interaction. For this reason, in Experiments 5 and 6 I include a condition where the sound continues but visual information is removed. If it is the case that Foulsham and Sanderson's participants were using the visual cue to guide their gaze, this manipulation should produce a very different pattern of gaze.

The findings also supported previous research by Tice and Henetz (2011) which indicated when auditory information is not present, more time is spent looking at targets who are listening and not speaking. A further interesting finding is that there was greater attentional synchrony in clips which included sound. Additionally the eyes of targets

continued to dominate fixations, regardless of condition contradictory to previous findings by Vo et al. (2012).

Overall, it is apparent that the manipulation of audio presentation and absence affects the way in which we follow gaze in static laboratory-based studies and also in more natural conversation with dynamic pre-recorded clips. Future studies should attempt to develop knowledge of the effects of audio when engaging in natural conversation and explore the link between and the amount we can possibly relate these findings in laboratory-based studies to real world scenarios.

Using visual cues to guide visual attention

In terms of which visual cues guide visual attention during a social conversation, there are many physical features of a person to think about.

First, the target who is present. It is well established that we look more to social stimuli and hence people within a scene. There is a strong attentional priority for social information (End & Gamer, 2017) and even more so when the dyad pair are interacting (Skripkauskaite, Mihai & Koldewyn, 2021). Whether this person is speaking, hence visually moving their mouth and gesturing may play a role. Physical, dynamic movement has also been shown to be more attractive than static imagery (Itti, 2005). Equally there may be physical features of the target which may mean our visual attention to them is increased. For example, if we find them attractive, if they have higher status or there is something abnormal about their appearance. The following literature discusses these visual features and how we use them as a cue to guide our attention during conversation.

For years, many saliency models have been created in an attempt to predict our viewing behaviour of a scene. However, the extent to which saliency models can be applied to images which include social stimuli often results in less reliable predictions. Cerf et al.

(2008) explored how saliency models performed within social scenes. They found that combining saliency model with a face detection model outperformed the saliency only model. Furthermore, this is supported in work by Birmingham and Kingstone (2009) who assessed the role of saliency on fixations by observers to social scenes. Their findings indicated that saliency did not account for the bias to eyes and neither to visual attention in general. Therefore, even if the eyes are not very salient in terms of bottom-up features, the eyes still attract a large proportion of our attention.

One reason for this could be that social scenes tend to involve more interesting and dynamic movement, with dynamic images being more visually attractive than static images. Itti (2005) established, using model prediction, that motion and temporal change were stronger predictors of eye movement behaviour compared to the colour, intensity and orientation of features.

Additionally, the element of surprise has been shown to attract our visual attention. Itti and Baldi (2009), explored how 'Bayesian Surprise' attracts human attention with 72% of all participants moving their gaze towards a location which is surprising. When you think about this in lay terms, you of course would be attracted to something which is 'out of the norm' or not as expected. For example, if you saw a street artist hovering in the air, we tend to have a curious compulsion to look. This perhaps is related to an evolutionary adaptation to ensure we are prepared and pay attention to elements of environment which could be a threat to our safety.

When thinking about engaging in a live conversation there are many visual features which come into play when deciding where to direct our gaze. Rather than solely relying on physical attributes (low-level), interactions are more complex, and the use of top-down information may serve to guide our attention more.

Foulsham et al. (2010) explored the role of dominance in gaze allocation. Their study involved third-party observers watching pre-recorded clips of individuals engaging in a decision-making task. The task involved a hypothetical survival situation such as where the targets had to decide which items they would need if they were abandoned on the moon. This is a task often used in group interview settings and quickly establishes who is the more influential or powerful members of the group. After taking part in the conversation, the targets were asked to rate the social status and influence of each target they completed the task with. When analysing the data from the third-party observers, unsurprisingly, the majority of fixations landed on the targets (77%). Due to the social nature of the videos and the dynamic movements from the targets, this is to be expected. The interesting finding comes from how the observers distributed their attention to the three targets present. The targets who had been classified as 'high status' received the majority of fixations, followed by the 'medium' status targets and the least fixations landed on those who were classed as 'low' status. The difference between the three levels of status was significant and reliable. This pattern was the same for the sum of attention throughout the clips (duration of fixations) and also for mean gaze duration. This demonstrates a predisposition to look to someone of higher status and raises questions as to which visual cues project this.

A consideration which is highlighted by Foulsham et al. (2010) is that perhaps those of higher status were those which spoke more. When analysing the distribution of attention over time with the verbalizations (which target was speaking and when), there was a significant effect of status on speaking time, with high status individuals speaking the most. However, the authors suggest as there was no significant difference in speaking time between high and medium targets, the attentional differences cannot be solely explained by the amount of speaking the targets did. Therefore, it appears that during a conversation, the perceived dominance of a target will directly influence the amount of visual attention they

receive. Questions then arise as to which physical features or gestures make a person appear more dominant.

During an interaction, individuals may not only grasp our attention, but they may also guide or divert our attention to other areas of the scene. This is most apparent when we think about magicians, who implicitly move our attention away from where they are deceiving us. Kuhn, Tatler and Cole (2009) found that the magicians gaze helped to conceal the misleading event, if their gaze was misdirected, demonstrating how participants followed the gaze of the magician. Participants looked less to the magician's hand if the magician's gaze supported this.

Combining audio and visual cues

Attempting to assess sound and vision together, and their role in visual attention distribution, becomes more complicated. However, it is rare that we experience one modality without the other in everyday social situations, with our attention often being guided by the fusion of the two. As such, the research to date tends to imply that the two together result in strengthening our ability to communicate. Perhaps, this is because this is due to familiarity – i.e., we generally are able to use both sound and vision in real world situations, but also perhaps there is a benefit to the use of both senses.

When thinking about social situations and the main exploration of this thesis, we must consider the effect of audiovisual integration on social visual attention during conversation. When taking part in a conversation with someone, we tend to look at their face to not only facilitate information processing, but also to indicate that we are listening. In such situations we tend to process both in parallel. There is research to suggest that manipulating the audio and visual inputs can affect our visual attention allocation. For example, different areas of the

face are attended to differently with the presence of audio and visual information (Vo et al., 2012).

Hirvenkari et al. (2013) explored audiovisual manipulations and how a non-involved observer's gaze is affected by the natural signalling displayed during turn-taking behaviours of two individuals. Their study involved asking participants to observe a pre-recorded conversation between a dyad pair, with no further instructions tracked. The clips shown to participants included the manipulation of audio (silent) and visual (freeze-framed) clip conditions, to explore which modalities evoked gaze shifting. Their findings were that both visual and auditory information (when presented in solidarity) generated shifts in gaze towards the speaking person. However, when both modalities were accessible, the gaze shift was significantly greater and faster, demonstrating the benefits of presenting both audio and visual information.

The results demonstrated that overall participants looked at the speaker on average 74% of the time. The authors argue that this percentage closely resembles that in a real-world dyad conversation, highlighting an observation by Argyle and Ingham (1972), where it was reported that a live listener looks at a speaker similarly 75% of the time. Hirvenkari et al. (2013) found that changes in the speaker directed the gaze behaviour of the third-party observers. These results indicate that the organization of turn-taking conversation has a strong influence on third-party gaze behaviour, rather than, for example, observers looking at speakers and listeners equally. This is not too surprising given evidence that we typically attend to the location of a sound source in various contexts. In a social situation there is also evidence to suggest that being able to view a speaker's face helps with perceptual ambiguity and conversation following. For example, Zion-Golumbic et al. (2013) explored the "Cocktail Party" problem whereby participants viewed conversational videos simultaneously with multiple sources of sound. Their results indicated that being able to view the speaker's

face enhanced the capacity for auditory cortex to track the temporal speech envelope of that speaker.

A comprehensive modelling approach, in a recent paper by Boccignone et al. (2020), presents a computational model which combines the audio and visual elements in an attempt to understand our deployment of gaze. They use the term 'patch' to describe the area in which a person is attending to, with a description of how 'foraging' the multimodal landscape is intertwined with the social context and sound. The authors state that a large amount of research effort has focused on salience estimation from natural scenes and often this research neglects the dynamics of actual attention deployment. They suggest that this is particularly apparent when gaze sampling is affected by goals, rewards, and expectations. Hence, in the authors proposed model, the 'foraging' dynamics are driven by audiovisual 'patches' which are able to change at any time depending upon the social value. The authors use both audio from the speaker as well as the knowledge that the eyes tend to fixate a speaker to create their best fitting model. By using model simulation experiments on social dynamic videos, the authors found there was an overall statistically significant similarity between their procedure and the scan paths of human observers. The authors do emphasise limitations including within-patch item handling such as facial expressions. Their research highlights that when exploring the two modalities together, the social context equally plays a role in gaze orientation.

Timing of looks

Gaze timing in video

However, it is not clear at which point observers move their gaze to a speaker during group interactions. Whether observers move their gaze in advance of a change in speaker or whether this gaze shift is reactive can be investigated by exploring the cues in conversation

which lead people to shift their gaze. In Experiment 5 I investigate the precise timing pattern and the presence of these cues, such as signals in speech or gestures and other physical behaviours. In order to understand these cues and their impact on gaze, we can manipulate the audio and visual content of the conversation.

For example, in Hirvenkari et al's (2013) previously described study, they found when looking at the temporal characteristics of gaze shifts, at the crucial turn taking transition, gaze predicted rather than followed speakership. Upon a turn-taking transfer, the attentional shift to the speaker was slightly before the beginning of the utterance, and, although alone both modalities evoked a shift, the anticipatory shift was most apparent when both audio and visual modalities were present.

A further example is Latif, Alsius and Munhall's (2018) study which investigated the role of auditory and visual cues on predicting turn-taking behaviour. Their study involved presenting participants with clips of a dyad pair engaged in a natural conversation. Participants were asked to watch the clips and respond with a button press when they felt the speaker was about to finish their turn of talking. The authors manipulated the stimuli by preparing trials where the audio or visual information was removed, giving three modality conditions: Visual-Only, Auditory-Only and Control. Decisions of turn taking behaviour were assessed with strongest performance when both audio and visual information was present in the Control condition. Participants responded significantly earlier in the Visual-Only condition in comparison to both auditory inclusive conditions. The authors deduce that visual information functions as an early signal indicating an upcoming exchange; whilst the auditory counterpart is used as information for individuals to precisely time a response to turn ends. This suggests that visual information might be critical for guiding gaze in advance of the next speaker, although anticipation was observed regardless of the modality presented.

These studies, which demonstrate the ability to predict speakership, involve dyad pairs. There is also evidence for anticipation in larger groups (e.g. Holler and Kendrick, 2015), with gaze moving before a change in speaker. However, the evidence in this case is more mixed. For example, a similar result to the aforementioned studies, demonstrating that gaze in third-party participants can predict speakership in a group setting, was reported by Foulsham et al. (2010). When analysing the temporal offset in the relationship between speaking and fixation, they found that participants tended to look to the speaker slightly (roughly 150ms) before the utterance beginning.

The temporal characteristics of following conversation were further investigated by Foulsham and Sanderson (2013). In this case, their study used a video-watching task and did not find that gaze moved in advance of the change in speaker. Instead, the authors reported that speaking preceded gaze by roughly 800ms, with the authors suggesting this may be due to the complexity of the video interactions.

In Experiment 5 I investigate the effect of visual and auditory modulations on the precise timing patterns of gaze to a speaker. While the role of visual information in the intelligibility of speech has often been studied, it is less clear how visual versus auditory cues are involved in the precise timing of gaze patterns. I thus include these manipulations to help understand the factors underlying how attention is deployed during complex social interactions.

Gaze timing in live settings

Gaze timing during conversation has also been investigated in live interactions, where both audio and visual signalling cues are available. Ho, Foulsham and Kingstone (2015), explored the precise timing of gaze during a live face-to-face conversation. This study provides evidence that an anticipatory effect occurs in a live conversation setting, with a lag

between changes in gaze and changes in speaker (roughly 400ms) similar to at least one study which used pre-recorded video (Foulsham et al., 2010). Ho et al. (2015) monitored dyad pairs engaging in two turn-taking games while both participants' eye movements were tracked. The authors assessed the temporal characteristics in terms of both gaze and participant-generated speech. This analysis enabled a detailed measurement of how speakers and listeners avert and direct their gaze. Interestingly, because this study looked at real people in a face-to-face situation, rather than someone watching a video clip, the results may reflect the dual function of social gaze (Risko, Richardson & Kingstone, 2016). In that, in a live environment, the eyes not only take in information, but they also signal to others. In other words, the gaze movements involved in live studies such as Ho et al. (2015) were not merely picking up on the information from the speaker but also sending a signal about listener engagement and turn-taking. This signalling can take place in real face to face interactions for example, when it is your time to speak; but not when looking at pictures of faces which are often used in classic social attention studies (Risko, et al., 2012). Considering the discrepancies about timing in the video studies described above, it is interesting to compare this behaviour between real and video conditions. If there are large differences between comparable "lab" and real interactions, then it would suggest that gaze to conversations is strongly affected by the ability to interact in the real situation. Arguably, these settings might show the same anticipatory effect for a different reason, in that the signalling cues that were exhibited in the live situation (which allowed for a dual function of gaze), may guide attention in the pre-recorded videos, hence allowing participants to pre-empt the speaker.

Role of head with gaze

In a natural conversation it is rare that eye gaze is alone without the use of head orientation and head movement. This physical movement which is generally seen in

conjunction with gaze is often observed simultaneously during a natural conversation and can act as a signalling cue.

Head orientation with gaze

Hietanen (1999) demonstrated with a response time study, that the visual information we extract from others gaze and head orientation is integrated. The study involved a presentation of a reaction signal preceded by a facial cue stimulus. The facial cue signal was either congruent, incongruent or neutral to the reaction signal. Findings were that the head stimuli of front and profile views with an averted gaze affected response times. This is in comparison to the frontal view of the face stimuli with a gaze looking straight forward. Furthermore, a profile view of the face with a congruent gaze cue did not result in the same effect. The author therefore concludes that visual information from others gaze is combined with head direction and the integration information appears to travel to the brain areas which produces visual attention orienting (Hietanen, 1999). Therefore, visual attention orienting appears to depend upon both eye and head orientation in a laboratory-based reaction task.

Bayliss, Pellegrino and Tipper (2004) produced a similar study to Hietanen (1999), by investigating the belief that the orientation of the head is the determinant which influences the gaze cue. Their study involved a digitized image of a face presented in the centre of a screen, with the face appearing in one of three orientations (upright, 90 degrees left and 90 degrees right). The eyes of the digital face could either be facing to the left or right of the eye or up or down in the conditions where the face was tilted. The results confirm the hypothesis that a vertical uninformative gaze cue could act as an attentional cue, if this cue is within a rotated face. Support for this can be found in research by Hietanen (1999) and Langten et al. (2000) whereby it is interpreted that eye gaze direction is influenced by head orientation.

Stiefelhagen and Zhu (2002) support this finding and highlight the importance of head orientation when detecting who is looking at whom. This study however expands upon more unrealistic lab work and is anomalous in that it examines participants in a 'real' interaction. The study involved four participants taking part in a group discussion set up to resemble a meeting. One participant wore head mounted equipment which enabled live tracking of the participants head and eye movement. The authors report that head orientation contributes to on average 68.9% of gaze direction. Furthermore, an estimate of where the target is focussing their attention can be accurately established on 88.7% on average using head orientation alone. These findings were based on a real-life scenario with four participants in a meeting room, which is a progressive leap forward in the quest to make social attention studies more social in nature. These findings are more representative of the naturalistic social interactions in everyday life. Consequently, future studies should expand upon this realistic research and take into account how head orientation attributes gaze behaviour. Although this thesis does not explicitly measure head orientation, I do use stimuli which include dynamic head movements.

Head movements with gaze

However, when we interact in social situations, we do not keep our head and eye movements stationary. Instead, we use our hands, face, head and body to signal to others using dynamic movements. It is therefore important to study more of the non-verbal dynamic cues that are present in everyday interaction.

Within Vo et al's (2012) study, which explored the effect of removing audio on gaze, they also took into account the effect of head movements when observing video. They found that during head movements, there was increased fixations on face regions. In particular, head movements led to an increase fixation on the nose which suggests that observers

adopted a centre bias during rapid movement to enable an optimal viewing position. However, this modulation affect was only found in natural clips where auditory information was available. Vo et al. (2012) suggest this finding shows that perhaps following a moving face may be particularly functional when following conversation. A further exploration of this could be to continue to analyse the eye movements of observers of natural group conversation which will include natural head movement and explore the same variables when the stimuli include multiple targets.

Since both head and eye movements contribute to our perception of where people are looking, in Experiment 5 I use videos where both cues are present, and in Chapter 5 (Experiments 7, 8 and 9) I manipulate the presence of the eyes which tests the uniqueness of this cue.

Atypical traits and eye-movements to conversation

Assessing eye-movements in healthy populations can significantly help us detect and even diagnose a range of disorders in which 'abnormal' eye movement behaviour is present as a key identifier. Exploring differing eye movements in such populations may not only help us to diagnose but also explore divergent cognitive processes in those that possess such traits. The next section explores eye movements in Autism (ASD) and attention deficit hyperactivity disorder (ADHD).

Social eye movements in ASD populations

When we think of abnormal eye movements, we often think of the inability to engage in direct eye contact at the correct moment, something which is often seen in individuals with autism spectrum disorder. The unusual orienting patterns have major clinical relevance as such individuals often socially interact in ineffective or inappropriate ways and often fail to

read social cues of their interlocutors (Benson & Fletcher-Watson, 2011). Studies have also demonstrated how individuals with ASD tend to show different and reduced attention to social stimuli, perhaps demonstrating a disinterest in socially relevant information (e.g., Riby & Hancock, 2009).

For example, children and adults with ASD often show abnormal and reduced social interactions, with atypical eye-contact, which has been established as predictor of ASD diagnosis (Baron-Cohen, 1995; Klin et al., 2002). In the lab, ASD individuals tend to not look at people in images and movies to the same degree as typically functioning participants (Dalton et al., 2005; Klin et al., 2002). In a developmental study by Falck-Ytter et al. (2013), the authors showed typically developing (TD) children and children diagnosed with ASD videos of two children interacting. The study explored the children's eye movements to the two target children on screen. The results indicated that when a target child gestured to the other target, the TD children tended to turn their attentions to the other target. The authors claim this is adaptive as this target has the power to decide what happens next. Contradictory to this, the ASD children showed a much weaker tendency to look towards the target upon the gesture. The authors suggest perhaps the ASD children fail to follow the course of events efficiently.

Klin et al. (2002), explored gaze in social situations in ASD individuals by assessing the time spent looking to people and objects in an eye tracking study. Their study used 5 digitised video clips each 30-60 seconds in length which were watched by 15 males with ASD and IQ-matched controls while their eye movements were recorded. The researchers analysed the fixation time on regions of interest including the eyes, mouth, body and objects. Findings suggested that increased autistic social impairment was correlated with more time spent on objects. Additionally, high social functioning scores were related to more fixations to the mouth area. Equally, there continues to be something special about the eyes, with authors

stating the best predictor of autism was reduced fixation times to the eye region of targets. Klin et al. (2002), suggest these results indicate how, when viewing natural social scenes, individuals with autism demonstrate atypical patterns of eye-movements.

Norbury et al. (2009), also explored fixation differences in ASD when watching a pre-recorded interaction. The paper used teenagers with ASD watching videos of peers engaging in familiar interactions. Again, results indicated that individuals with ASD spent less time looking to the eyes and were slower to fixate this region of interest.

Attempting to explain this, a meta-analysis by Chita-Tegmark (2016), suggested that there are two levels in which social attention in ASD is atypical. First, in ASD individuals, attending more to social stimuli than non-social stimuli (which TD's undoubtedly exhibit) is reduced. The second level explores the way attention to social images is deployed in terms of specific regions of a target (eyes, mouth, and body). In TD populations, the eyes appear to be of increased importance when distributing attention to social scenes. As is discussed throughout this thesis, we know the eyes are 'special' in typical populations, whereas ASD individuals tend to show increased attention to the mouth and body (over the eyes), (as discussed in Chita-Tegmark, 2016). These social atypical attention behaviours are however disputed, with some researchers finding no visual attention differences (e.g., Kemner et al. 2007).

Kuhn et al. (2010) explored the effect of these atypical visual attention differences using magic tricks. The authors proposed that perhaps the ASD population may be less susceptible to these tricks due to their inability to follow social cues like typical individuals. Hence, perhaps the increased attention to other regions of the body would result in them uncovering the magic. However, the opposite was found, in that autistic people were more vulnerable to the trick. The authors explained this in terms of ASD individuals presenting

with a difficulty of rapidly attending to all targets (both social and object), providing incongruous patterns of behaviour.

The author of the meta-analysis (Chita-Tegmark, 2016) explains that there are vast differences in experimental procedures and perhaps there are specific circumstances in which ASD populations show diminished attention to social targets. The analysis evoked that when studies used stimuli which had a high social content, social attention in ASD was most impacted. Chita-Tegmark (2016) suggests that future studies should identify which aspects of a social stimulus are informative in ASD, which contributes to the rationale for Experiment 7. In this experiment, we will use video clips of natural social conversation to address how individuals of high trait ASD visually follow conversation. To date, there are limited studies which investigate ASD with dynamic, real interactions, which Experiment 7 will address.

There are multiple suggested reasons as to why individuals with ASD respond to social settings differently. One suggestion is that the social cues which TD populations are attuned to aren't processed in the same way. There are further suggestions that perhaps ASD individuals do not look to this cue (in that it doesn't grasp their attention), or do not know to look at the cue, insinuating perhaps this was never learnt. Equivalently, it could be that this cue does still gain attention, but ASD individuals find this cue particularly aversive.

One argument, the 'dialectical misattunement hypothesis' explains the social difficulties are caused by a disturbance which happens at an interpersonal level, (Bolis et al., 2017). This is described as a misalignment of attuning to other people's social signals during an interaction. Therefore, this suggests perhaps it is not the individual who has deficient attention to the social stimulus, but the extraction and use of the social cues is not in line with the behaviour of a neurotypicals. In Experiment 7, I explore the ability of high trait ASD individuals to use social cues (the eyes) in gaze following.

In support, Freeth and Bugembe (2019), who used a face-to-face social interaction, found that opportunities for reciprocal social gaze were missed by adults with autism. This was due to the ASD individuals looking to the experimenter less than TD's when there was direct eye contact.

In comparison, Cañigueral, Ward and Hamilton (2021), found contradictory evidence to suggest high-functioning autistic individuals were able to use their gaze as social signals. Their recent study explored to what extent the effect of being watched modulated gaze behaviour. The study used a method which enables the researchers to adjust the level of real social engagement with three conditions (live face to face, video call and video). A key hypothesis was that ASD individuals would exhibit fewer looks to the confederate, in comparison to the typical groups. The authors hypothesised this would be apparent for all conditions and based this on previous findings by Von dem Hagen and Bright, (2017). Despite their hypothesis and the well-known lay knowledge that ASD individuals exhibit more of an aversion to eye-contact, there were no large differences between ASD and typical groups when analysing eye gaze patterns to eyes and mouth. Both groups were able to use both perceiving and signalling functions to plan gaze allocation equally when looking at speech and the time course analysis and both groups gazed more to the confederates when listening than when speaking. The authors claim this shows how autistic individuals are able to modulate their gaze behaviour during a conversation depending on their current role (speaking or listening). In fact, one of the only differences to note was that ASD individuals surprisingly gazed more to the eyes of the confederate, something which contradicts previous research and many lay people's real-world experiences. When investigating the effects of the three conditions, participants overall looked less to the eye of the confederate in the live and video call conditions, compared with the video clip condition (consistent with Laidlaw et al., 2011), with no differences in ASD. However, when including speech in this analysis,

individuals averted their gaze more when speaking, in all conditions. This could be seen as surprising given participants knew the video was pre-recorded (hence no ability for live interaction and no live signalling). This result is however congruent with findings in Experiment 5, where live and third-party eye movements were analogous. The authors suggest that averting the eyes while speaking could be due to cognitive demands.

Overall, the authors suggest high functioning ASD individuals do not exhibit gaze patterns which are atypical to that of a normal population, in that there is not a reduction of interest to the faces of others. Perhaps this is not apparent in high-functioning ASD and equally the authors note that this may not be true for spontaneous interactions. Despite this, the study makes a leap in ecological validity by being the first to systematically compare gaze in clinically diagnosed ASD when engaging in live and pre-recorded social interactions.

In a similar set up, Freeth, et al., (2013), did find differences in ASD traits and atypical groups when modulating social presence. The study measured the proportion of time spent viewing a confederate when face to face or on a pre-recorded video and correlated the viewing behaviour with traits of ASD. Their findings demonstrated that those with greater traits of ASD looked less to people when watching videos. However, interestingly there was no differences in the live situation. Freeth et al. (2013) suggest perhaps the increased attention to faces in video in populations with low ASD traits, is because another person's face and gaze are extremely captivating, regardless of the context. Therefore, it may be that those with high autistic traits do not experience the same captivation, with the video stimulus being less interesting to this population. Freeth et al's (2013) paradigm used a one-to-one conversation. In Experiment 7 I aim to explore this further by using stimuli of a larger group conversation.

It is important to note that in Freeth et al's (2013) study, a sample of the general population took part, and the differences were observed in those showing high traits of ASD,

rather than being clinically diagnosed as in Cañigueral, Ward and Hamilton's (2021) study. Cañigueral et al. explain that perhaps the reason there are no differences in their findings is due to the clinically diagnosed individuals being able to control and regulate their gaze in a live condition. Almost as if this is a well-practiced, learned behaviour from real interactions. Equally there is the need to consider the fact the individuals knew they were taking part in a testing session and perhaps in the live situations they are able to adjust their behaviours to reflect those of TD's, in a sense, acting. This would support the reason for no differences in the live condition in Freeth et al's (2013) work.

To underpin the source of this difference and highlighting the atypical and spontaneous eye movements in ASD, Jiang. Kreigstein and Jiang (2020), explored the under-researched brain mechanisms involved in such gaze patterns. Using brain imaging techniques, they uncovered that the level of Autistic traits presented (measured on the Austism Spectrum Quotient (AQ)) could be predicted by activity in the posterior superior temporal sulcus (pStS) and its connectivity with the fusiform face area (ffA) when engaging in eye contact with an interlocutor. Hence, perhaps the brain mechanisms involved in eye contact with others could help us to predict autism in individuals.

Overall, despite common beliefs that ASD individuals exhibit 'abnormal' eye contact, the considerable scientific research on the topic is less conclusive. In ASD individuals, there seems to be some differences in social attention. However, these differences only seem to emerge under certain conditions and the differences are not clear cut.

For this reason, Experiment 7 questions to what extent is third-party attention to conversation is affected by traits of ASD and whether there are any attention differences when we occlude the eyes, hence removing information of a key social cue. A second disorder of interest, and one which is often comorbid, is ADHD. Research regarding ADHD and social attention abnormalities is limited, which I will discuss this next.

Social eye movements in ADHD populations

In terms of the effect of ADHD on social gaze behaviours, there is limited evidence. However, when looking at general oculomotor behaviours, there is fairly conclusive evidence that there are frontostriatal deficits in ADHD individuals which cause inhibitory eye movement differences. This section explores this research and ends by explaining how this research could inform how social gaze may be affected.

ADHD is one of the most common mental disorders in children. There is an average of around a 5% prevalence globally. However, it is still relatively underdiagnosed, in particular in girls and older children (Sayal et al., 2018). The DSM-5 (American Psychiatric Association, 2013) defines ADHD as a persistent pattern of inattention or hyperactivity which interferes with the individual's development or functioning. This could be presented in the way an individual fails to give their full attention to schoolwork or other activities. Those with the disorder may appear to not be listening when spoken to and easily distracted. The hyperactivity element can be demonstrated in an inability to sit still, often fidgeting and a sense of restlessness. In a social situation, patients may talk excessively and have trouble waiting their turn to speak. ADHD is often difficult to diagnose as a number of other mental disorders can have similar symptoms such as: sleep disorders, anxiety and certain types of learning disabilities (CDC, 2020). There is no one agreed psychological test, biomarker or neurophysiological assessment for ADHD, with symptoms presenting differently at different ages.

With the symptoms displaying erratic and inattentive behaviour, this disorder can be difficult to study from a visual attention perspective. At first glance, we may assume the individual may present with eye movements which mimic their attention, perhaps unpredictable and inconsistent with typical populations. Consistently, a frontostriatal pathophysiology has been suggested to be the cause of the symptoms of ADHD which leads

to a reduced ability to inhibit behavioural responses (Munoz et al., 2003). Hence, oculomotor tasks can be used to probe the ability of such populations to inhibit reflexive responses. The understanding being that those with ADHD will have impaired or atypical oculomotor behaviours compared to control conditions.

Research has started to explore eye movements in ADHD developmental populations. This method of data collection has an advantage of being a passive experience for the participant, not requiring any cognitive skills, yet providing us with extremely rich and detailed data (De Silva et al., 2019). For this reason, this method can be easily performed on children at the age at which ADHD is first diagnosed. Despite the task being simple for a participant, testing such populations can be difficult for a researcher. For example, due to the nature of the ADHD symptoms, there can be difficulties in the data collection process (e.g., calibration and focussing during the testing session).

In line with the idea that ADHD may be exhibited in a general deficit in oculomotor control, Hanisch et al. (2006) support this notion in their study which tested eye movements to measure specific aspects of oculomotor behaviours. They found that patients, in comparison to controls, were specifically impaired in stopping an already initiated response and suppressing exploratory saccades during novel situations. Their data support the notion that there is an underlying impairment in cognitive inhibition (associated with prefrontal lobe functions) which can be seen explicitly in such eye movement tasks.

In an additional study which explored various saccade tasks, Mostofksy et al. (2001), state their findings from oculomotor measures support the idea that ADHD patients have deficits in prefrontal functions (in particular response inhibition). The authors suggest this deficit contributes to the atypical behaviours observed in ADHD children.

Specifically looking at female children (an often-underdiagnosed population) with clinically diagnosed ADHD, Castellanos et al. (2000), found that in ‘go-no go’ tasks, patients

made three times as many intrusion errors (that is saccades in the absence of a ‘go’ or ‘no-go’ stimuli) compared with controls. The authors suggest this data confirms that girls with ADHD have an impairment of executive function.

In addition to saccade inhibition differences, children with ADHD have also been shown to have an elevated response time and delayed initiations of serial search during visual search tasks (Karatekin and Asarnow, 1998), as well as inaccuracies of visuo-spatial working memory tasks (Rommelse et al., 2008). Furthermore, there is often a predisposition to spend significantly more time gazing to irrelevant areas on continuous performance tests (Lev et al., 2020), together with inconsistent and atypical scanpaths during reading (Mohammadhasani et al., 2020).

When exploring if there is any effect of ADHD on social attention, Serrano, Owens, and Hallowell, (2018) compared the eye movement behaviour of an ADHD and a control group in an emotion identification task. The authors used images with seven different facial expressions. They found that participants with ADHD spent less time looking at the social area and specific areas such as eyes, and mouth, in comparison to the control group. In addition, the ADHD group had slower reaction times than the control group when asked to identify the emotions. These findings suggest that ADHD individuals have atypical movement behaviour especially observing social stimuli compared to controls.

Research had also been conducted to help understand how reading may be impaired in children with ADHD. Deans et al. (2010) found those with an ADHD diagnosis displayed significantly shorter fixations and a lower proportion of left to right saccades compared to a control group. The study also highlighted how a large number of participants were excluded from the study due to excessive head and body movements, a general problem associated with collecting data from such population. This is one of the many aspects to consider when testing atypical populations and additionally when combined with testing developmental

populations. As also highlighted in this paper, some of the symptoms of ADHD are similar to other mental disorders. With ADHD being high in comorbidity it is often difficult to determine whether any differences found were due to the child's ADHD diagnosis.

In relation to this, an additional consideration during data collection is the patients use of medication. A commonly used medication in the treatment of ADHD, methylphenidate, has been shown to affect saccades, errors and reaction times during anti-saccade tasks (Klein, Fischer, & Hartnegg, 2002). For this reason, in Experiment 8, I ask participants to disclose if they are taking any medication for ADHD symptoms.

It is apparent there are differences in oculomotor behaviours in children with ADHD. Whether these differences are recorded due to complications in data collection, or whether there are true differences found, has some potential for controversy. Overall, it seems that the eye movements of ADHD patients, which could be affected by the pathophysiology's of the prefrontal cortex, display a different pattern to the typical population in terms of exploratory and volitional saccades, precise oculomotor control and inhibition functioning (Huang & Chan, 2020). Despite this, to my knowledge, there is limited research on how such eye movements affect social situations in ADHD patients. One suggestion is that such population may show a different pattern of fixations when looking to a social stimulus. A key symptom is that those with ADHD often seem uninterested in conversation and appear to not be listening or distracted. Therefore, it could be hypothesised that participants may have less control of their eye movements to follow a 'common' social gaze pattern and instead their eye movements may present more erratically or appear to be without purpose. Experiment 8 explores the effect of ADHD traits on social attention while participants observe a dynamic group conversation.

Other contributing factors

There are many other additional variables which are worth merit when discussing attention to people within social conversations. Here, this section touches on aspects of the target individual such as: social status, attraction, gestures and emotions as well as individual differences of a participant. All of which have research to suggest they may affect how visual attention is distributed.

Social status

When thinking about how attention is directed within a group setting, we can think of how our own attention is shared when engaging in a conversation with multiple people. In a group conversation, often the person who asserts the most dominance within that interaction will receive the most attention (Foulsham et al., 2010). This can be explained in terms of an evolutionary approach, with an increased attention towards these individuals perhaps enabling other lower status members of the group to learn from their leader, (Maner, DeWall & Gailliot, 2008).

In a study which advances on static imagery to investigate gaze to third-party videos of conversation, Foulsham et al. (2010) explored the extent to which perceived social dominance affects gaze. They used pre-recorded videos of people taking part in an interactive conversation whereby the needed to make a decision. The social status of each individual was rated by the people taking part in the clips. Foulsham et al's results suggested that the perceived status of the peer ratings acted as a predictor of where the third-party participants looked, with higher status people looked at more often and for longer.

Attraction

In addition to this, there is an abundance of research which has found that our attention is directed towards something (or someone) we find attractive (Maner et al., 2003). Again, from an evolutionary perspective, because of their inclination of better reproductive health, we know certain phenotypical features of faces are generally perceived as more attractive (Fink & Penton-Voak, 2002). The level of perceived attraction may also influence our visual attention during a conversation. There is evidence to suggest that observing a dynamic, moving image of individuals may affect reporting of physical attractiveness. Riggio et al. (1991), asked judges to rate physical attraction and likeability of individuals from videotaped and photographed interactions. Findings indicated that the type of stimulus modulated attraction ratings. The authors argue this taps into the multifaceted nature of attractiveness and how we must include variables of dynamic attractiveness (such as nonverbal expressive behaviour) in future research. Therefore, perhaps during conversation (which is dynamic and fluid), we cannot predict that the most facially attractive individuals will be gazed at most often and we should also include other visual cues such as dynamic expressive displays.

Gestures

When looking at the effect of gestures on attention during social interactions, Gullberg and Holmqvist (2006), found that only a minority of gestures (8.8%) drew fixations and that the face continues to dominate as a fixation target. Langten et al. (2000) suggest that when someone is perceived to be directing their attention, eye gaze is not the only cue to the sensitivity of shifts of eye movement. For example, they suggest that other areas we must not neglect include cues such as orientation of the head, posture of the body and gestures. As it is

assumed that these cues are processed automatically, they should all be analysed in future research in reference to decisions on social attention (Langten et al., 2000).

Despite this, in a recent mobile eye tracking study, Kajopoulos et al. (2021) established participants fixated mostly on the face of the experimenter. This was despite pointing gestures and directional gaze movements of the experimenter. This adds further evidence of the strong drive to attend the face during social interactions.

Scott, Batten and Kuhn, (2019) demonstrated how the action being carried out can affect the rate of fixations to the face and other body parts. Their study involved tracking participants eyes while they watched three types of social interactions (monologue, manual activity and active attentional misdirection). The results indicated that during a condition which the actor presented a monologue, as expected, most time was spent looking at the face. In the other conditions which involved activities, the gaze location changed to relevant body parts. The authors deduce that humans are able to use a strategy and top-down processes to influence and control their attentional focus.

Emotion

When analysing social attention in conversation, other factors such as the mental states of the targets may play a crucial role in the way our attention is dispersed. The role of emotion on gaze following has been researched by Ohlsen et al. (2013). Their study involved priming individuals with either a threat or no threat condition, with the attempt to induce either a dangerous or a safe environment. Their findings demonstrated that the emotional context significantly influenced the gaze cueing effect. This research suggests that certain gaze behaviours are influenced in an emotional context. The authors propose this suggests an implicit, context-dependent follower bias which carries implications for a wide range of research into social cognition and visual attention.

Moreover, Buchan, Pare and Munhall (2007) researched the role of emotion on eye movements during social tasks. Their study involved participants watching pre-recorded clips of an actor saying 27 sentences expressing happy, neutral and angry emotions. One of the researcher's objectives was to assess if, when viewing different emotions, different gaze fixation patterns would be observed. Their results showed no statistical difference between the three emotions with the fixation distributions being extremely similar against the three emotions analysed. The authors do however report a difference in gaze shift when analysing data between tasks. When judging emotions, participants preferentially shifted their gaze to the eyes more often than when asked to recognise speech. This therefore suggests the task in hand is a determinant of the eye movement when watching emotional clips.

Furthermore, the effect of task on eye movement has been considered with the role of emotion suggested as a future direction highlighted in Foulsham and Sanderson's (2013) paper. Their projected research suggestions include that gaze may be different if targets' emotion was evaluated by participants, supporting the idea that perhaps different gaze patterns directed to the eyes and mouth may be present with different tasks or stimuli. Their suggestion is that perhaps if the participants were to evaluate emotional expressions, gaze may be more skewed towards the mouth. This line of research would further the understanding as to whether gaze during conversation is controlled by motion cues and aspects of the stimulus, versus whether the gaze is under strategic control of the observer.

This thesis does not focus on the role of emotion in context as to how this affects fixation to speakers, instead a conscious effort was made to ensure there were not large differences in emotion intensity within the stimuli.

Inter-individual differences

When examining gaze behaviours of the participant themselves, evidence has explored the importance of accounting for individual differences in race (Crosby, Monin & Richardson, 2008), age (Frank et al., 2009) gender (Shen & Itti, 2012), social status (Foulsham et al., 2010) and culture (Chua, Boland & Nisbett, 2005). The above-mentioned research demonstrates the population differences which should be considered when exploring oculomotor behaviours.

At an individual level, it is also important to note that even personality differences may affect gaze behaviours, with curiosity significantly affecting scene viewing (Risko et al., 2012). In regard to social attention research specifically, it may be that some participants are simply better at understanding social scenarios and hence may follow a different pattern of eye movement. For example, it has been shown that subjects, who are trained to follow specific scanning patterns, continue to show individual differences in eye movements (Keppel & Wickens, 2004). The individual differences in performance are just one of the challenges involved in eye tracking research.

In addition, Chapter 5 of this thesis explores the effect of a further inter-individual difference, traits of disorders, and their effect on visual attention to conversation in more detail.

Culture

Another individual difference highlighted is the cultural differences of eye movement when analysing scene perception. Furthermore, this can be a criticism of Western research in that findings are only representative of a particular population in which they were reported. Cultural differences in gaze behaviour have been explored by Chua et al. (2005), in a scene perception study. The study aimed to establish the differences between Chinese and American participants when confronted with a naturalistic scene. Findings were that Chinese

participants focused on the focal object less and slower than the American participants. This supports the research which suggests Western cultures attend more to focal objects, whilst East Asians attend more to contextual information. This cultural difference in scene perception extends to a social context, with McCarthy et al's (2008) study which assessed how culture and context modulates gaze display. Their study demonstrated how social factors driven by participant culture affect gaze behaviour. Comparing Japanese and Canadian participants, in face-to-face question tasks, Canadian participants tended to look up when thinking and when they were aware they were being observed. Japanese participants however, looked down when thinking, even when they knew they were being observed. McCarthy et al's (2008) study demonstrated how thinking-related gaze behaviours are present in such situations and how this varies between cultures.

Moreover, in a cross-cultural mobile eye tracking study, Haensel, Smith and Senju (2021), established crucial differences in West Caucasians and East Asians, with East Asian dyads spending more time engaging in mutual gaze. The authors suggest this challenges gaze avoidance as an observation observed cross culturally.

The research in this thesis was conducted solely in the UK & Canada and does not explore the effect of culture in conversation. However, this may be an important next step which could be pursued.

Chapter Summary

This introductory literature review has highlighted how far the field has progressed in exploring social interactions with the use of visual attention. I have demonstrated findings which explain what attracts our visual attention and specifically the looking behaviours presented within social interactions. Finally, I have touched on which factors influence fixations to speakers and gaze locations in previous experiments. The following research in this thesis expands upon previous findings of group interactions and uses stimuli with a more complex and dynamic methodology, which explores the gaps highlighted in the current literature. In the next Chapter, I explore the methods used in eye tracking research.

Chapter 2: Methodological challenges of eye tracking

The three experiments (Experiment 1, 2 and 3) presented in this chapter form part of a publication ('Theory of mind affects the interpretation of another person's focus of attention') published in Scientific Reports by Dawson, Kingstone and Foulsham, 2021.

Chapter 2 explores the techniques used when collecting visual attention data. First, I explore the use of static and mobile eye-trackers (MET). Then, I present three experiments exploring the perils of coding cursors which resemble MET data. Here, interestingly, theory of mind interplays with a simple decision of where a cursor is located.

Methodological history

The methods in which eye movements are recorded have dramatically improved with advances in technology. Early methods included: eye movements by visual observation, the use of magnifying glasses, microscopes, reflected beams of light, affixing a mirror to the eye, recording the corneal blind spot, still picture photography, and electrooculography (Yarbus, 1967). Often the methods were intrusive, inaccurate, and sometimes uncomfortable for the participant, often requiring anaesthesia to the eyeball.

Advances in methodology

Advances in eye tracking methods have enabled the recording of eye movement on a moment-by-moment basis with increasing accuracy. Using non-invasive techniques, we can now track the eye using only a camera and a short calibration procedure. The data we are able to extract gives us insight into what captures the observers attention, what the observer chooses to focus on and provides us with clues as to how stimuli are perceived and the cognitive mechanisms involved (Duchowski, 2017). As methods advance, the spatio-temporal accuracy is now more precise which enables recent research to pinpoint exact moments in which various oculomotor behaviours occur. This in turn, helps us to uncover determinants of visual attention and human cognition.

When attempting to study where we look, eye tracking can give us a detailed and real-time insight into our visual attention which arguably can be difficult to report consciously. The advantages of using eye tracking methods includes the ability for us to reveal natural and subconscious behaviour in a high level of detail. The eye-tracker removes the need for a researcher to subjectively make assumptions about attention and the data produced is often more objective with quantifiable information which can be easily compared across participants. Upon analysing eye tracking data, it is important to note that the data gives us an

insight to the visual system and in turn we make inferences about the cognitive counterparts involved.

The subsequent studies within this thesis use a combination of static and mobile eye-trackers for data collection. For the purpose of this thesis, I refer to static eye-trackers as desk-mounted equipment with pre-recorded stimuli where a participant is restrained and mobile eye-trackers as a wearable device where a live participant is free to move around the scene. The choice of which is carefully considered, depending upon the studies aims.

Desk mounted eye-trackers

Eye-trackers are a real-time technique to measure the gaze location of the eye on a moment-to-moment basis. For static, desk mounted eye-trackers, the equipment generally involves an infrared light which is projected onto the eye, whilst a camera records how the light is reflected off of the cornea. Static eye-trackers are more commonly used when we want the participant to visually explore a pre-recorded stimulus which is presented on a screen, rather than the natural world around them.

Head movement

One problem associated with most static eye-trackers, is the need for participants to keep still whilst the data is collected. This often involves participants placing their head into a chin and/or forehead rest for the duration of the study. This set up requires participants to preferably not move their head, to obtain the most accurate data. The problems associated with eye tracking; in particular the natural movements of the head, have been developed and researched extensively (Nguyen et al., 2010). Such eye tracking techniques tend to have a low tolerance for head movement as users need to hold their heads still, which feels unnatural

(Zhu & Ji, 2007). This is particularly apparent when testing infants or patients with disorders which mean keeping still is more taxing. Despite this, methods have been developed to allow unrestrained head movement when using static eye trackers. For example, using a sticker target placed on the participant's forehead, such as that developed by SR research.

Niehorster et al. (2018), explored the variability in data loss across five of the most popular eye-trackers which allowed data collection involving a non-optimal and unrestrained pose by the participants. Niehorster and colleagues, asked participants to perform six tasks which were designed to investigate 1) how the eye-trackers coped having lost the eyes, 2) how the eye-trackers responded when only one eye was available and 3) the level of performance when participants presented with non-optimal head orientations. The study was designed in mind with the belief that often a manufacturers' claims of robustness do not match with user-experience in infants. Interestingly, the results indicated that the Eyelink in particular had trouble when participant's head orientation was not at an optimal position. This study sheds light on cautions when using eye tracking with unrestrained participants. For this reason, for experiments in this thesis which use static eye tracking, participants will be restrained with a head and chin rest.

Wearable mobile eye tracking

Mobile eye tracking builds upon the method of using desk-mounted static eye-trackers allowing free movement and locomotion. As well as collecting all of the valuable visual attention information in real time, this method aids the study of live situations with the addition of a built-in scene camera and microphone. This enables exploration of the participant's environment as well as the ability for the researcher to examine the participants visual perspective on a moment-to-moment basis. This technique often requires participants

to wear a pair of glasses which include the infrared cameras and a scene camera on the front of the glasses (see Figure 2.1).

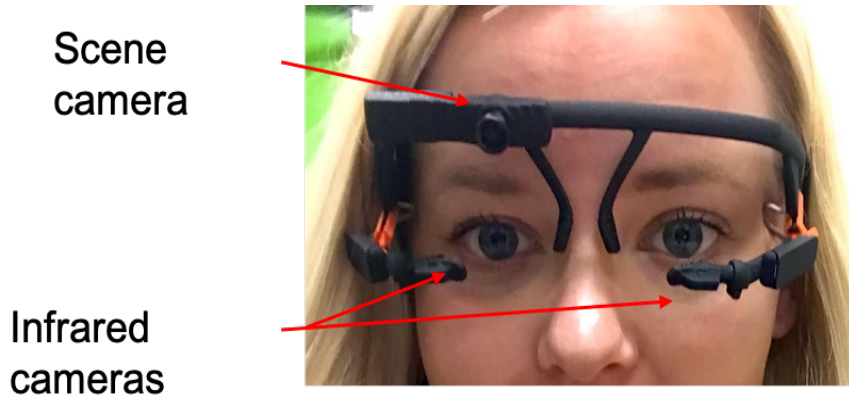

Figure 2.1. Image demonstrating an example of wearing a mobile eye-tracker (Pupil Labs), with scene camera and infrared cameras identified.

The data collected therefore allows you to map eye position onto a live scene of the participant's moving environment as seen in Figure 2.2.

Figure 2.2. Image of a real mobile eye tracking experiment which depicts the participant's view from the scene camera and a fixation point (blue circle) to show where within the scene the participant is looking.

This type of data collection has many benefits, including the ability to use the eye tracking glasses in mobile, dynamic environments, gaining large ecological validity when comparing them to a pre-recorded lab-based alternative. The flexibility to use this method in most environments allows an abundance of new research opportunities to record visual attention. This is a popular method and provides a great insight for market research and consumer behaviour; for example, how people explore a supermarket visually when in locomotion.

A question the novice psychologist may ask is 'why do we not use the wearable glasses for all eye tracking research?' considering they appear to have many benefits which overcome restrictive problems of the static method.

Problems with dynamic eye-tracking

Wearable mobile eye-trackers

With wearable eye-trackers, the lack of restriction does have many positives but collecting this data has more methodological problems.

One example is how the glasses draw attention from the general public, often making the participant feel self-conscious, which in turn could affect their behaviours. Depending on the model, the glasses may also become uncomfortable and sometimes even limit the participant's peripheral view, with many participants finding stairs difficult to navigate when wearing the device. In turn, the glasses may slip or be adjusted by the participant, which can result in significant errors and large increases in gaze deviation (Niehorster et al., 2020). Equally, researchers should also be aware of changes in the environment, including lighting and the weather, which can often leave data unobtainable.

In a recent paper, Hessels et al. (2020) verbalises and examines two common misconceptions of wearable eye trackers. First, that they are cost-effective. This is true in

terms of the equipment purchase, however, there is extensive manually coding required in the analysis hence making the time commitment unproductive. The authors should also consider the need to be cautious in referring to the data as 'real world' and possessing greater ecological validity. They state that often researchers do not provide an explanation of why the characteristics they are measuring could not be elicited in a laboratory setting. Second, they highlight problems with the data itself. Not only are the terms "fixations" and "saccades" ill-defined and often ambiguous when coding which leads to data which is incomparable, but also the equipment itself may perhaps not be as reliable in detecting gaze direction as we would optimistically hope for. Experiment 4 within this thesis offer an authentic account of this. For this reason, there are pros and cons to static and wearable eye trackers which should be considered when designing the study.

Dynamic areas of interest

One of the main problems with eye-tracking data lie in the area of interest (AOI) analysis; this is true for both wearable and static eye-trackers which use moving interest areas. Often, when collecting data from a wearable device a researcher or their assistant must painstakingly manually code the data. Depending upon the research objectives, this may include analysing each frame to code where the fixation point resides. For example, is the fixation on a person or a target, and at what time and for how long? This can be an extremely long process. This leaves a lot of room for human error, combined with extremely subjective analysis. (See Experiments within this chapter for a more detailed account of coding subjectivity). What counts as a 'hit' (described as a fixation on or off a target) could vary between researchers, with large inter-rater variability with the potential for coders to subconsciously adopt the perspective of others. An additional analysis problem is comparing across conditions and between participants is very difficult. This is because participants will not necessarily (and rarely) be looking at the same thing at the same time, as they move

around the scene. Equally, participants may not even look to a target that you are hoping to assess. Hence, the stimuli may not always be present in the scene camera, unlike pre-recorded methods, where it is possible to present the same stimulus on screen at the same time across all participants. Despite this, if using pre-recorded stimuli, if there is a moving AOI there is still room for human error. For example, in Experiments 5-8, I use dynamic AOIs (e.g., to locate a targets eye or mouth position). These moving areas are small and require the experimenter to draw these boxes and map the movement onto a video stimulus, which is equally as meticulous. Although the same AOIs are used for each participant, and are drawn prior to data analysis, there is still an element of subjectivity.

The following experiments (1-3) investigate the of analysing eye tracking data, which was personally experienced during data analysis.

Experiment 1, 2 and 3: Is this a hit? Theory of mind affects face bias when coding mobile eye tracking data

Aspects of the research are taken from a publication ('Theory of mind affects the interpretation of another person's focus of attention' in Scientific Reports by Dawson, Kingstone and Foulsham, 2021). The data analysis is the same as the published version, with additional results reported in the Appendix (1).

Mobile eye tracking (MET) can provide us with extensive and rich data from a natural setting, something which is invaluable in social attention research. However, analysing the data often requires extensive manual coding which can be subjective. These experiments were prepared to implicitly study the subjectivity of coding mobile eye tracking with a view to explore how animacy and ToM can affect coder decisions. In three experiments, we investigated how descriptions of a cursor affect how a novice person codes a 'hit' (on target) when making judgements about the location of a cursor in a scene. In Experiment 1, participants were told that this cursor represented the gaze of an observer and were asked to decide whether the observer was looking at a target object. This task is very similar to that carried out by researchers manually coding eye tracking data. In Experiments 2 and 3, we explored whether this bias occurs when removing information about biodata and instead told the participants the cursor reflected a 'random' computer system, or a computer system designed to seek targets. Overall, it appears the ability to adopt the perspective of observers interplays with what should be an objective decision.

Introduction to Experiment 1, 2 and 3

From shortly after birth, humans are drawn to animate and biological elements of the environment (Sifre et al., 2018). Indeed, across the lifespan, human attention tends to prioritize animate beings, such as humans and other animals, over inanimate items. This is reflected in dissociations between the representation and processing of biological animate and inanimate items in the brain (Naselaris, Stansbury & Gallant, 2012; Grossman & Blake, 2001) and a behavioural bias toward animate items (Kovic, Plunkett & Westermann, 2009; Pratt et al., 2010), both of which may confer a number of evolutionary advantages (New, Cosmides & Tooby, 2007).

Gaze, Perspective Taking and ToM

One specific instance of this preferential bias for biologically relevant stimuli can be found in the human tendency to select and follow the eye gaze of other conspecifics (Friesen & Kingstone, 1998; Emery, 2000; Frischen, Bayliss, & Tipper, 2007), which has been linked extensively to theory of mind (ToM), (Baron-Cohen, 1995; Nuku & Bekkering, 2008; Teufel et al., 2009; Foulsham & Lock, 2014). ToM describes the cognitive capacities which underlie our understanding of other people. These are often measured by asking people to judge what other people know, or why they behave the way they do, and there is considerable scientific interest in understanding how ToM develops and how it is related to behaviours such as perspective taking and empathy (Samson et al., 2010; Singer & Tusche, 2014). Visual perspective taking, the ability to understand what other people see, gives us useful information during social interactions (Baron-Cohen, 1995) and is a critical building block of ToM. In other words, "putting yourself in another person's shoes" by adopting the visual perspective of another person. For this to occur, the individual must be aware that their self-perspective differs to another's perspective and that an object can be perceived differently

depending upon the perspective adopted (Apperly & Butterfill, 2009). Measuring the impact that another person's gaze direction has on an observer's attention has frequently been studied (see Chapter 1 for more details). Recent work, however, suggests that gaze following may not require nor measure ToM (Cole et al., 2016; Kingstone et al., 2019), see Cole and Millett (2019) for a review. The present study therefore takes an alternative approach and turns the traditional gaze following approach on its head, by measuring whether ToM affects the interpretation of another person's gaze direction.

MET data

We achieve this goal by exploring to what extent this preference for faces over objects is present when asking novice people to make judgements about the location of cursors. We do this by modifying the cursor description, with participants in Experiment 1 believing the cursor to be mobile eye tracking data. Mobile eye tracking data builds upon static, desk mounted eye-trackers, allowing free movement for the participant to explore a live scene, not only with their eyes but also their head direction. As stated, despite the vast ecological benefits of this method and the flexibility of use, an ongoing concern of choosing this technique lies in the analysis stage. Often a researcher has to painstakingly hand code the data, manually mapping gaze to an object (Niehorster, Hessels & Benjamins, 2020). Depending upon the research objectives, this may include analysing each frame to code where the fixation point is. For example, is the fixation on a person or a target, and at what time and for how long? In a recent paper, Hessells et al. (2020), highlight how this isn't cost effective with 10 minutes of recording taking multiple hours to code. This leaves a lot of room for human error combined with subjectivity.

Solutions to this problem have been suggested with recommendations to use computer programmes to remove the human error of manually coding data. For example, software has

been advocated, where, after assigning areas of interest around the face, an algorithm is able to locate the AOI and confirm or refute if the fixation has landed on it.

Brône, Oben and Goedemé (2011) explore to what extent this type of semi-automatic coding is effective within complex settings in the wild. Their account claims object recognition algorithms may aid the process of analysing real-world behaviour in eye tracking. Advocates of such automatic processes, such as Hessels et al. (2019), highlight why researchers are not frequently using automatic AOI-construction methods. They highlight that the methods are technically complex and, to date, not effective enough for empirical research.

The present experiments were designed upon experiencing this subjective coding experience first-hand. Equally, from previous experience working with other coders when analysing fixations in MET data, it was noted there was an increased subjective assumption. That was, when coding social situations, coders often assumed the participant *must* have been looking to the face of a person within the scene.

Present research

In three experiments, we present observers with prototypical data collected from a mobile eye tracking study and ask observers to indicate if the fixation cursor, which represents the gaze direction of another person (in Experiment 1), is directed toward different items (objects and faces) in a visual scene. This is a task that researchers may have to complete when coding such data, but the possible impact of perspective taking, ToM and one's goals on those decisions has not yet been investigated.

Although some studies have shown that participants can make judgements about another person's intention by looking at their eye movements as represented by a fixation cursor (Foulsham & Lock, 2014), it is also the case that we are surprisingly unaware of our own fixations (Foulsham & Kingstone, 2013; Clarke et al., 2017; Kok et al., 2017). In the present study we can test whether knowledge of what people are likely to look at (the

animacy bias) and the ability to adopt their perspective, can be applied to a fixation cursor. Across three experiments observers are told that the position of the fixation cursor is generated from MET data from a human (i.e., one who does have a mind and goals, Experiment 1), randomly by a computer (i.e., one who does not have a mind or goals, Experiment 2) and by a computer vision system (i.e., an agent which does not have a mind but does have explicit goals, Experiment 3).

Additionally, in Experiment 1, we include 4 different types of cursors, to understand whether size and shape affects participants decision.

As humans, unlike computers, are preferentially biased toward animate items in the environment, we predict that observers would be biased to report that a fixation cursor was directed to an animate item versus an inanimate object only when the cursor was understood to be generated by a human. However, if we find similar results throughout the studies, with scenes including faces more likely to be coded as on target than objects throughout, this will imply that there is an overarching predisposition and bias in how the coder represents faces and objects.

Any differences between studies should be considered when exploring the perils of MET analysis. If we assume participants will be completely objective, as per computer automated AOI analysis, we expect participants to code the cursor as on target *only* if the two are interconnected. Any subjective differences between Experiment 1 and Experiment 2/3 will be interesting to compare with a reference as to why changing the story behind the cursor, from computer generated to biodata, modifies participant decision subjectivity.

Experiment 1

Method

All experiments were approved by the Ethics committees of the University of British Columbia or the University of Essex, and all research was performed in accordance with institutional guidelines. Informed consent was obtained from all participants. Experiments were pre-registered.

Participants

426 (321 female) volunteers were recruited online and via posters at the University of Essex and the University of British Columbia.

Stimuli

Drawing from staged scenes taken on a university campus, we selected 10 animate scenes each containing a different person, and 10 inanimate scenes each containing a different object. Each image measured 930×671 pixels. Onto each scene we placed a red cursor (that differed in shape or size: a large or small circle or cross). These cursor types were selected to explore whether different shapes or sizes of cursor, which are commonly used with eye tracking data, affected decisions regarding eye movement behaviour. Each of these cursors could occupy one of five different distances from the target object, with the nearest cursor at the edge of the target, and the distances increasing horizontally (left or right) in steps of 15 pixels (Figure 2.3), with the vertical position fixed. In images of people, the faces were in profile with the cursor always placed to the front of the face. Collectively, 20 scenes (10 animate, 10 inanimate) x 4 cursor types x 5 distances yielded a set of 400 images for this study.

Figure 2.3. Experimental stimuli. The left panels provide an example of a small circle cursor whose centre is displaced 15 pixels (Distance 2) from the nearest edge of a person ((A) animate scene) or object ((C) inanimate scene). The right panels provide an example of a large cross cursor displaced at a maximum distance of 60 pixels (Distance 5) for a person ((B) animate scene) or object ((D inanimate scene).
We have full informed consent from the individual depicted for the publication of this image.

Design

Participants were randomly assigned to one of the cursor shapes (between-subjects). The within-subject factors were target type (person or object) and cursor distance (5 distances). Each participant saw 20 of the 100 possible images for their cursor condition, randomly selected with the provision that each original image (10 animate and 10 inanimate) was presented only once but all 5 distances were represented for both target types.

Procedure

Participants judged cursor location via an online survey (Qualtrics). After reading the instructions, participants were provided with an explanation of eye tracking and shown an example video clip of a cursor representing eye gaze moving around a scene. Participants were instructed that researchers have to make decisions as to whether the person was looking at an object of interest (a "hit") or not, and that where they were looking was depicted by the cursor. Participants were made aware of the subjectivity of gaze cursor coding decisions, given some inaccuracies that could be seen in the video clip. It was explained to participants that researchers have to code whether a cursor is on the target, a 'hit', or not by deciding whether the cursor is on target. More specifically, participant instructions were: 'For the purposes of this research, pretend you are a researcher analysing eye tracking footage. In a moment, you will be shown 20 still images from a live video recording. You will then need to decide if the Focus Point is a 'hit' (on target) or not.'. Following this, participants were asked 'Is this a 'hit'? and given the name of the potential target ('ball', etc), for each of the 20 images. Participants selected 'Yes' or 'No' before the next image was presented in a randomized order.

Results

These studies were pre-registered and data, scripts and methodological details are available online (https://doi.org/10.17605/OSF.IO/NEM6B).

We analysed the relative frequency of "hit" judgements for objects and faces, split by the five levels of Distance (1-5) and by Cursor Shape and Size. For descriptives see Appendix 1. We used a generalised linear mixed model (GLMM) approach, using 4 predictor variables (Distance, Target Type, Cursor Size, and Cursor Shape) to predict the binary response and thus in which circumstances participants would classify the cursor as a hit. Each

participant (426) responded to each image (20), giving 8520 data points. We used the lme4 package in R and a binomial function, assessing the contribution of each factor with maximum likelihood. Participant and scene were included as random effects. Where possible we also included random slopes by participant and item, and these were dropped when models failed to converge.

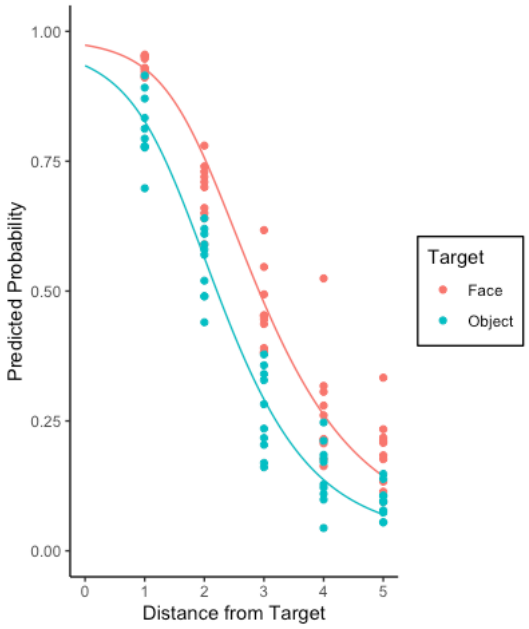

Figure 2.4 The likelihood that a participant will code a cursor as a hit in Experiment 1 for a face or inanimate object. Lines show the average marginal probabilities estimated by GLMM. Data points show observed probabilities for each particular scene.

Figure 2.4 shows the empirical data and the best fitting statistical model. The continuous variable of Distance was a significant predictor (compared to intercept-only: $\chi^2_{(3)} = 1027.8$ $p<.001$). As expected, the probability of a cursor being coded as hitting the target decreased as distance from the target increased ($\beta = -1.96, \pm 0.08$ *SE*, $p<.001$). Adding Target Type (object or face) further improved the model ($\chi^2_{(4)} = 206.12, p<.001$). There was an increased probability of reporting a hit when the cursor was near a face, compared to when it was near an object ($\beta = -1.36, \pm 0.10$ *SE*, $p<.001$). In additional models, we added Cursor

Size and Shape, but these did not improve the model fit (p=0.43 and p=0.19, respectively). Thus, the size and shape of the cursor did not make a difference to whether a participant would code the cursor as a hit. The interaction between Distance and Target Type also failed to improve the model fit (p=0.93). Table 1.1 gives full details of the best fitting model, which includes random effects of participant and image and random slopes of Distance and Target Type by participant.

Fixed effects	Estimate	SE	Z	p
Intercept	5.85	0.24	24.85	<.001
Distance	-2.05	0.08	-24.41	<.001
Target Type *(face/object)*	-1.36	0.10	-12.97	<.001

Table 1.1. The best fitting GLMM for predicting the binary decision of cursor location in Experiment 1. The reference level for Target Type was the face condition.

As illustrated in Figure 2.4, as distance away from the target increases by 1 step (15 pixels), the hit rate drops by roughly 20%. However, participants treated selection of faces and objects differently. If the target was a face, the predicted chance of a hit was 10-15% higher than when the target was an inanimate object. This difference was fairly consistent across the 5 distances measured.

Collectively, these results show a clear difference in the way that the location of a gaze cursor relative to a target is evaluated based on whether the target is a face or an inanimate object. When the target is a face, participants are more likely to judge that the cursor indicates that a human observer is looking at the face than when the target is an inanimate object. This outcome provides support for our hypothesis that observers will be preferentially biased to report that a fixation cursor is directed to an animate item versus an inanimate object when the cursor is understood to be generated by a human. Interestingly, for

both types of target, there is a graded response, indicating that participants did not only consider the cursor as selecting the target when it fell on or near to the target. Even when gaze (i.e., the cursor) was some distance away, participants were willing to interpret it as reflecting attention to the target, especially when the target was a face.

It is tempting to attribute these effects to the judges' theory of mind. By this account, the cursor is more readily judged to be targeting a face because the judge attributes a mind to the looker, and they know that such an observer is biased towards animate objects. However, an alternative possibility is that because the judges themselves are humans with minds, it is their own attention that is being pulled toward the animate items in the scenes (supporting work by Pratt et al, 2010). This would explain the marked tendency to report that the cursor is directed toward the target when it is a face rather than an inanimate object.

To distinguish between these two explanations, we conducted a second experiment, the key change being that participants were told that the cursor was randomly generated by a computer. This should remove any preconceived beliefs about the attributes of the looker from whom the cursor was generated. If Experiment 1's results reflect attributions of the mind to the looker, which is represented by the cursor, then in Experiment 2 the preferential bias to report the cursor as directed towards faces (rather than inanimate objects) should be eliminated. However, if the results of Experiment 1 reflect the judge's (i.e., the participant's) own attention being drawn toward the faces, we should find the same results as before and regardless of the instructions. Of course, these two explanations are not mutually exclusive, and the results of the current experiment may reflect both the participant's attribution of mind to the looker and their own attentional bias, in which case one would expect the preferential bias to report that the cursor is being directed toward to faces may be reduced but not eliminated.

Experiment 2

As cursor size and shape did not matter in Experiment 1, we ran only one cursor condition in Experiment 2.

Material and methods

Participants

An additional 100 (39 female) volunteers were recruited online via prolific.ac.uk. This sample size is approximately the number of participants in each of the cursor conditions in Experiment 1. We also ran a power simulation (using the 'devtools' package (Kumle, Vo & Draschkow, 2020)) to confirm this size of sample would give us excellent power to detect differences between face and object (>95%).

Stimuli and design

The same 20 images from Experiment 1 were used, with 5 levels of distance, resulting in 100 images with the same cursor type (a small circle). The factors of target type and distance were manipulated fully within-subjects as in Experiment 1.

Procedure

Participants completed the same task as in Experiment 1, with the only difference being the instructions given beforehand. Rather than being given information and instructions about mobile eye tracking, participants were told that the position of the cursor was "randomly generated by a computer". It was explained to participants that they would be asked to help code whether a cursor is on the target. Participant instructions stated: 'We want to know whether the cursor has randomly landed on an object/person. If it has, we call this a 'hit'. Please respond to each image with 'Yes' or 'No' if the cursor is a 'hit'. They were then asked, precisely as before, to code the images by indicating whether the cursor reflected a 'hit' (in other words whether it was 'on' the target in the scene) or not.

Results

Figure 2.5 shows the overall percentage of 'Yes' responses/'hits' for each condition. For descriptives see Appendix 1. For statistical analysis we again used a GLMM with random effects of participant and scene to fit the binary response variable, fitting 2000 data points (100 participants responding to each of 20 images). We began by also including random slopes, but this often led to models which could not be estimated. We therefore omit random slopes, although in cases where such a model did converge the outcomes were the same.

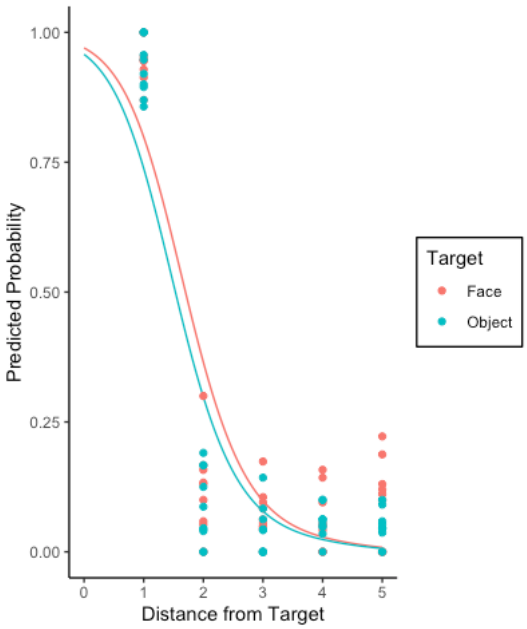

Figure 2.5. The likelihood that a participant will code the cursor as a hit in Experiment 2 for a face or inanimate object. The lines show the average marginal probabilities estimated by GLMM and the scattered points indicate the observed probability for each image.

First, we added the continuous variable Distance as a predictor to our intercept only model. This improved the model ($\chi^2_{(1)}$ =990.56 p<.001), and the probability of a cursor being coded as hitting the target decreased as distance from the target increased (β = -2.16, ± 0.12*SE*, p<.001).

However, including Target Type did not result in a significantly improved model, ($\chi^2(1) = 3.69$, p=.055). Therefore, in Experiment 2, whether the target was an object or face did not affect cursor location judgements. This finding disconfirms the hypothesis that the results in Experiment 1 reflect the observers own attentional biases and supports the hypothesis that in Experiment 1 their preferential bias to report that a cursor was directed toward faces reflects their attributions of mind to the looker.

Comparing Figures 2.4 and 2.5, it is clear that responses in this experiment were quite different. Here, there was a large decrease in hit responses from Distance 2 onwards. As distance away from the target increases from step 1 to 2 (15 pixels), the hit rate drops by roughly 80%. After this, the rate of positive responses remains low and fairly constant. This indicates that, while participants tolerate some distance between the cursor and the target when it comes from a human, when it is generated by computer they do not. There are minimal differences between objects and faces, although a slight tendency for more 'Yes' responses to faces at larger distances.

The key finding from Experiment 2 is that when the fixation cursor is described as being randomly generated by a computer, participants judge the location of a cursor the same, whether it is positioned near an animate item or an inanimate item in the visual scene. In particular, there was no difference between classification of face and object trials at the nearest distance, and when the cursor was further away it was rarely endorsed as landing on the target. This does not mean that the judges' own attention was never biased towards the animate items, and this may account for the slightly more frequent responses in face trials at further distances, but such incidences were rare.

Although it is clear that the change in instructions affected judgements, the attribution of a mind to the source of the cursor may not be the only explanation for this difference. In Experiment 1, the observers were told that the images reflected biodata from a human, and

we argued that judges were reasoning about the mind of that human (for example intuiting an animacy bias). In Experiment 2, we suggest that these ToM processes were not applied when the cursor was generated by a computer. However, this experiment also removed any sense of a goal, with cursors explained as 'randomly generated'. It is possible that observers avoided classifying hits as there was no particular reason for the "computer" to select anything. In Experiment 3, we refined the instructions, explaining that the cursor was a computer vision algorithm designed to seek and detect targets, and making the instructions as similar as possible in all other respects to those in Experiment 1. If behaviour in this case is the same as in Experiment 1, it would suggest that having a goal rather than representing a mind is the critical factor.

Experiment 3

Material and methods

Participants

A further 100 (32 female) volunteers were recruited online via prolific.ac.uk.

Stimuli and Design

The same 20 images from Experiment 1 and 2 were used, with 5 levels of distance, resulting in 100 images with the same cursor type (a small circle). The factors of target type and distance were manipulated fully within-subjects as in the other experiments.

Procedure

Participants judged cursor location via an online survey (Qualtrics). After reading the instructions, participants were provided with an explanation of a computer vision system designed to seek and detect targets within a scene (the targets being faces and other objects). Participants were shown the same example video clip that we showed in Experiment 1, but this time they were told it reflected the selections of the computer system.

Participants were given instructions reflecting that of Experiment 1 and asked to help code the images, given some inaccuracies that can be seen in the video clip. Instructions were: 'For the purposes of this research, please help us to determine whether the computer system has successfully located the target. In a moment, you will be shown 20 still images from a live video recording. You will then need to decide if the computer cursor is a 'hit' (on target) or not.', Following this, just as in the prior two experiments, participants were asked 'Is this a 'hit'? along with the relevant target label, for all 20 images. Participants selected 'Yes' or 'No' before the next image was presented in a randomized order.

Results

For descriptive statistics see Appendix 1. Figure 2.6 shows the overall percentage of 'Yes' responses/'hits' for each condition. We used the same statistical GLMM analysis, again fitting 2000 data points (a further 100 participant responses to 20 images).

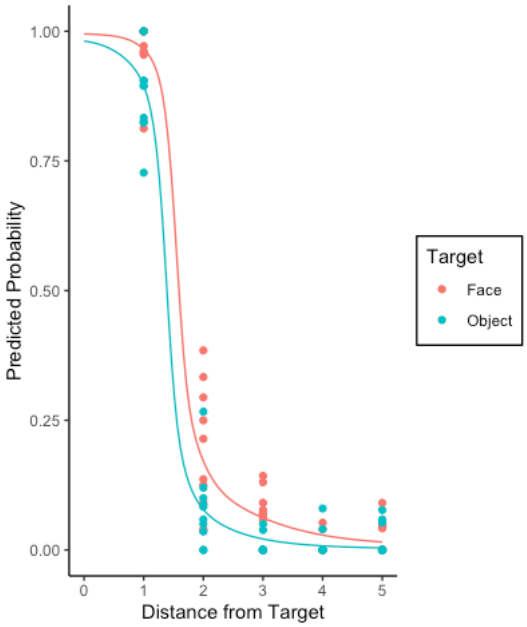

Figure 2.6. The likelihood that a participant will code the cursor as a hit in Experiment 3 for a face or inanimate object. The lines show the average marginal probabilities estimated by GLMM and the scattered points indicate the observed probability for each image.

We followed the same analysis steps as the prior two experiments. Optimal models included random effects of participant and item and the random slope of Distance by participant. First, we added the continuous variable Distance to our intercept only model, which, as before, significantly improved the model, ($\chi^2_{(3)}$ =1561.7 p<.001). Again, the probability of a cursor being coded as hitting the target decreased as distance from the target increased (β = -10.42, ± 1.80*SE*, p<.001).

Second, we added Target Type, which resulted in a significant improvement of the model, ($\chi^2_{(1)}$ =15.74 p<.001). There was an increased probability of reporting a 'hit' when the cursor was near a face, compared to when it was near an object (β = -1.90, ± 0.43 *SE,* p<.001). Comparing Figures 2.4 and 2.6, the current experiment produced a difference between faces and objects at some distances, but this was less pronounced than in Experiment 1. We also observed an improvement when we included the interaction of Target Type and Distance, but this only occurred when random slopes were omitted, ($\chi^2_{(1)}$ =5.38, p=.02). Differences between faces and objects were only noticeable at Distance levels 2 and 3, a similar trend to that observed in Experiment 2.

Between experiment analysis

In order to compare the effect of changing participant instructions in more detail, we performed a comparison between experiments. We combined the data into the same model, comparing Experiment 1 (where judges were told the cursor was human gaze), Experiment 2 (where cursor position was "randomly computer generated") and Experiment 3 (where the cursor represented a computer vision algorithm).

To confirm there were no effects of sample size differences, we ran this analysis with only participants who saw a small circle in Experiment 1 (1/4 of the total sample size) matching Experiments 2 and 3. The between experiment analysis, with this adjusted sample, produced the same significant results as an analysis with the full sample size. Results from the adjusted sample are reported below (for a version of Figure 2.4 based on this restricted sample, see our data repository: https://osf.io/nem6b/).

We combined the data from the three experiments into one model. Our baseline model with Participant (308) and Image (20) as random effects gave us 6,160 data points.

We then added the three predictors Distance, Target Type, and Experiment (1, 2 or 3) in separate models building on the last. All stages provided significant improvements on the previous model and all factors were significant. In addition, we observed interactions between Target Type, Distance and Experiment, demonstrating that differences in responding between face and object varied across the experiments. To examine this in more detail, we ran separate comparison analyses using the same model building approach.

First, we compared Experiments 1 and 2. We again added the significant predictor variables of Distance and Target Type. Adding Experiment into the model also significantly improved the model, ($\chi^2_{(1)} = 41.15$ $p<.001$). Then, we added the interaction of Target Type and Experiment. Model comparisons demonstrated a significant improvement, ($\chi2(1) = 21.48$ $p<.001$) and a reliable interaction ($\beta = 0.88, \pm 0.19$ *SE,* $p<.001$). This confirms that the effect of Target Type was different, and negligible, when participants rated the cursor believing it to represent random selections by a computer and not human gaze. In a final model we also added interactions with Distance and the three-way interaction ($\chi2(3) = 66.93$ $p<.001$). This model indicated that the effect of distance was also different in Experiment 2. The three-way interaction with distance may indicate that the bias seen with ToM in Experiment 1 is more apparent at some distances (when compared to Experiment 2).

Comparing Experiment 1 with Experiment 3 led to similar results. Adding the 3 predictor variables significantly improved the fit of the model. In particular, adding Experiment as a predictor variable, resulted in a significant improvement on the model, ($\chi^2_{(1)} = 40.05$ $p<.001$) Adding the interaction of Target Type and Experiment demonstrated a further improvement, ($\chi^2_{(1)} = 8.00$ $p = 0.0047$). Including interactions with Distance was beneficial for model fit ($\chi2(3) = 201.1$ $p<.001$), but in this case the three-way interaction was not reliable.

Using the same analysis to compare Experiment 2 and 3, adding Distance and Target significantly improved the models. However, when adding Experiment to the model, this did not result in a significant improvement ($\chi^2_{(1)}$ =0.60 p=0.44) and there was not a significant effect of Experiment (β = -0.21, ± 0.28 *SE*, p=0.45), demonstrating no significant differences in cursor coding between the two experiments.

To confirm the differences between experiments, we ran an additional analysis to quantify bias to faces. For each participant, we calculated an average 'bias score', measured by subtracting the average frequency of positive responses to objects, from the average frequency of positive responses to faces, pooled across all distances (see Figure 2.7). Between subjects t-tests indicated that a face bias was significantly higher in Experiment 1 than in Experiment 2 (t (206) =3.43, p=.001) or 3 (t (204) =2.97, p=.003). Experiments 2 and 3 were not significantly different (p=.54). Together with the GLMM results, this analysis gives strong evidence that Experiment 1 involves different judgements.

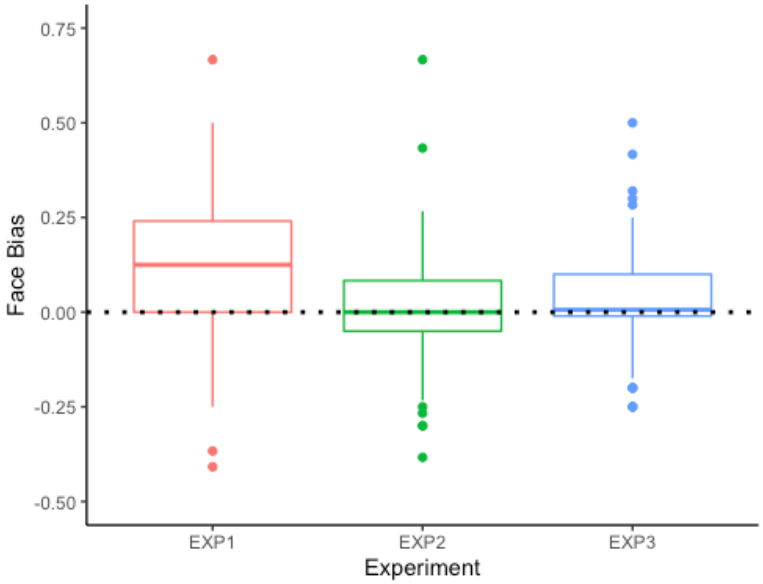

Figure 2.7. Boxplot to show the average face bias for each experiment. A score of zero (dotted line) indicates the participants judged objects and faces equally. Positive scores indicate a bias towards faces. Boxes show the median and quartiles with outliers represented as dots beyond.

Collectively, these results confirm that changing the believed source of the cursor changes the way that it is judged with respect to animate and inanimate objects. We suggest, in Experiment 1, participants are applying what they implicitly know about ToM to the gaze location of another human, which affects the interpretation of an otherwise non-social stimulus. When participants believe the cursor is randomly computer generated (with no preconceived ideas of agency, goal, or ToM) there is no distinction between animate and inanimate targets. When participants believe the cursor is generated by a computer which is seeking out targets, they show a slight bias to endorsing hits near a face (compared to other objects), though otherwise they behave similarly compared to when judging a "random" computer. There remains a significant difference between the pattern observed in Experiment 1 and Experiments 2 and 3. In contrast, and as can be seen in Figures 2.4, 2.5 and 2.6, the differences between Experiments 2 and 3 are minor and non-significant.

Discussion of Experiment 1, 2 and 3

Previous research has indicated that humans are biased to attend to animate objects. It has also been argued that we follow gaze in an automatic way which reflects ToM, an interpretation that has recently been criticised (Cole & Millett, 2019). The present study took an altogether different approach, asking people to make an explicit judgement about the location of a cursor which – in Experiment 1—reflected the gaze of a third party (MET data). We reasoned that if observers attributed mental states to the cursor, perhaps adopting the perspective or considering ToM of the observer whose attention is represented, they would be more likely to interpret the cursor as selecting an animate object than an inanimate object. If

this behaviour results from attributing mental states to the "looker", then it should not be exhibited when the cursor is controlled by a computer. This was tested in Experiments 2 and 3.

The results reveal effects of animacy and agency on an apparently simple judgement: deciding whether a cursor is selecting an object in a scene. In Experiment 1, participants were more likely to code the cursor as 'on target' when it was near to a face as opposed to an inanimate object. In Experiment 2, when participants believed that the cursor represented selections made randomly by a computer, and hence the computer not having a ToM perspective per se, the pronounced difference between faces and objects was eliminated. Comparisons between the two experiments demonstrates that there is an underlying predisposition to believe people are looking at animate objects and shows how judgement decisions are affected by knowledge of others' intentions. In Experiment 3, when participants believed that the computer selections of the items in the scene were goal-directed, a bias towards judging a cursor as being directed towards faces was detected. However, this bias was markedly smaller than in Experiment 1, and failed to yield a significant effect when compared against Experiment 2. The increase in judgements to faces may reflect both a bias of the coder's own attention towards faces (present in all experiments) and an effect of attaching a mind to the cursor (present only in Experiment 1).

A strong implication of this research is that in Experiment 1, participants were able to adopt the perspective of the human observer, which resulted in participants inflicting their own animacy biases onto their decisions. Arguably, in Experiment 2, this effect was eradicated as participants could not adopt the perspective of a 'random computer' selection. However, interestingly, in Experiment 3, participants did show a slight animacy bias when the observer was described as a computer with goal to 'seek' the target. This research relates

to work such as Samson et al. (2010), where participants were able to adopt the perspective of a non-human avatar.

The task in the present studies, to decide whether a cursor is on a particular target, appears to be simple. However, our results indicate that the same distances are treated differently if the target is a face, and that the same task yields very different judgements when the cursor is believed to represent human rather than computer behaviour. We believe this could be a powerful paradigm for measuring ToM and perspective taking. Given its simplicity, versions of this task could be useful for these measurements in children and those with developmental disorders. For example, although individuals with autism typically show impairments in social attention, they can make judgements about the gaze of others, at least in some conditions (Morgan, Foulsham & Freeth, 2020). The current paradigm could provide a way to probe the perspective-taking component of such judgements. Our experiments also mimic the judgement that is made when researchers manually code eye tracking data (as is frequently the case with mobile eye-trackers). The implication is that this manual coding may be biased towards social targets in a way which has not been previously investigated.

In Experiment 3, we used instructions that were closely matched to those in Experiment 1 by describing a computer vision system that had a goal to select faces and objects. On the one hand, this experiment produced some of the same animacy bias we saw in Experiment 1. This indicates that, even when the cursor is generated by an artificial agent, if that agent has a goal, participants may reason about this goal (a ToM-like process). This could involve introducing biases that participants expect to be present, such as assuming that like themselves the computer system is going to have a bias toward animate objects, as that is a common aim in human-designed computer vision systems (e.g., face-recognition systems).

On the other hand, the general pattern of responses in Experiment 3 was more similar to Experiment 2. In both of these experiments, there was strong agreement that the cursor was

on the target when it was only very close to the face/object. This was quite different from the more graded response seen in Experiment 1. In that experiment, even as the cursor moved away from the target, participants showed a marked tendency to identify it as landing on the target, especially if the target was a face (e.g., 20-25% of the time judging a cursor as landing on a face at the two most extreme distances). It is also possible that the mere presence of a face in the scene increases responses to anything, a general social facilitation which could be tested by using images with both a face and a non-face target presented in parallel. However, this cannot explain the way that the same cursor is treated differently when it is associated with human gaze.

In terms of the methodological concerns, from these findings, we must consider how certain factors may affect the perils of manually coding MET data. The results of the experiments highlight how important it is to include reliability rules. It seems coding a fixation cursor as on versus off target it not as objective as we would hope. This suggests it is important to implement coding reliability rules amongst coders, and furthermore, report these rules in publication. Positively, the size and shape of the cursor did not affect coder decisions, suggesting this is not an important factor to consider when setting up the eye-tracker.

To further understand this gaze interpretation bias, this simple paradigm (with small stimuli changes) could be used to explore a range of factors which may affect coder biases.

One example is exploring the presence of this effect with different scenes and targets. For example, is it that faces are really special, or would we see this effect in anything biological or any body part? For instance, would this effect be present if using the target individual's torso opposed to the face? Based on this studies results, we would suggest there would be an effect, but perhaps less prevalent. Additionally, is it necessary that the face is present, or would the back of the head also produce this result? We may assume that, if our

results are due to the participant adopting the idea that the cursor is biodata and hence must be relevant to our own visual biases, we would expect the back of the head and torso to have less of an effect. Additionally, if our own visual biases come into play, would we also see this result in extremely salient objects in an environment? Our results suggest the human aspect over an object has more weighting in this effect.

Furthermore, it would be interesting to test the strength of the bias to faces. For example, it would be interesting to use scenes where the faces are in less obvious locations, perhaps not the focal point of the scene. In the above study, we chose images where the targets (both objects and faces) were fairly central to the image, to replicate that of MET data whereby the participant tends to orient their head (and hence the scene camera) to their target of focus. Hence it is common for the observer to place where their visual attention currently is, in the middle of their gaze environment by moving their head. However, it would be interesting to explore the effect of manipulating this. Would coders assume that the observer must be looking at target that is central to the scene? Or would faces (e.g., to the far-left background of the scene) continue to dominate preference?

A further contemplation is the effect of affordance on an object and the orientation of the face. In this present study, the cursor was placed in line with the target features (e.g., in the direction of the facial orientation or near the handle of an object). Future work should explore different angles and cursor locations in relation to the target features.

An additional consideration could be differing levels of agency and experience in the described observer. For example, would people be more lenient if we told them the eye tracking data and hence cursor was generated from a baby, robot or a dog? In other words, would the ability to adopt an observer's perspective interplay with this ToM element?

Conclusions

Overall, these experiments have examined to what extent participants adopt others' perspectives and impose their own biases during simple judgements of locations. When making these judgements, participants are preferentially biased towards judging that people are looking at animate targets such as faces rather than inanimate objects. Critically, this bias is weak or eliminated when the exact same cursor is associated with a nonhuman source that has only a goal or no goal at all, respectively. The strong implication is that participants are making a theory-of-mind type attribution process to the human cursor (i.e., an animacy bias) that is not made for the computer cursor. Thus, the present study demonstrates that observers attach minds to representations of other people's gaze, not only to images of their eyes, but also to stimuli that represent where those eyes are directed, and this attribution of mind influences what observers believe an individual is looking at. Hence, this study has uncovered the need to be more vigilant when coding MET data as well as explored how faces are special even in this (what should be objective) context.

Chapter Summary

The present chapter has explored the methods used in social visual attention research including the advances, the benefits and also the perils.

The above experiments explored to what extent changing the story behind a cursor affects biases towards faces when making a decision about the location of the cursor. It is clear there are distinct differences between Experiment 1 and Experiment 2, which may be caused by an ability to adopt another's perspective and ToM. In Experiment 1, we replicated a common situation in mobile eye tracking research, whereby researchers have to manually code fixation points from their moving data. Participant's responses indicated a tendency to believe the gaze was on a person more so than an object, at all cursor locations. As expected, as distance away from the target increased, the frequency in which participants coded the fixation as a 'hit' decreased. Cursor size and shape did not affect the 'hit' rate.

To confirm this finding was not an effect of the stimuli used, in Experiment 2 and 3, we tested the role of ToM, by presenting the same scenes to new participants but now changing the story, with the statement that the cursor was generated by a 'random' computer system or by a computer system designed to seek targets. The bias to report that the cursor was directed toward faces was abolished in Experiment 2 and minimised in Experiment 3. Hence indicating ToM interplays with a simple decision regarding the location of a cursor. As highlighted in these experiments, researchers should be aware of this coder bias when manually coding MET data.

The remaining chapters of this thesis explore a range of social interaction questions using both static and wearable eye tracking methodologies, (with careful consideration to the methods chosen and conclusions drawn).

Chapter 3: Visual attention in live interactions

The following chapter includes a live mobile eye-tracking experiment involving a live social interaction. Two main manipulations are included: eye contact and group size. These two factors are explored to assess their effect on eye movements.

Experiment 4– Look into my eyes: the effect of group size and eye contact in live conversation

As human beings we continuously interact with one another in group settings. For example, in an office meeting, in school or around the dinner table with our families. Previous research has demonstrated the sophisticated dance of eye movements that we display to facilitate such interaction (e.g., Ho et al., 2015). However, little is known about how we visually engage in naturalistic *group* conversations (that meaning engaging with more than one other individual). This exploratory mobile eye tracking study investigates two factors: the effect of group size and the effect of eye contact on visual behaviour in live settings.

More specifically, the experiment within this chapter explores the effect of removing eye contact signalling during a live interactive conversation. Here, I refer to eye contact as looking to another person. Using a mobile eye-tracker, I investigate participants' eye movements during a natural conversation with a researcher (and confederate). The researcher subtly manipulates their eye-movements to either A) reflect a normal gaze in conversation and B) stop making eye contact with the participant. Additionally, the effect of group size is measured by directly comparing eye-movements in conversations of groups of two and three.

Introduction to Experiment 4

Present Study

This exploratory experiment aims to investigate the effect of group size and the influence of eye contact on looking behaviour when interlocutors are engaged in a live social interaction conversation. Participants will engage in a fluid conversation and form a dyad and a triad. As this study is exploratory research (and taking into account methodological factors discussed in Chapter 2), I will assess the overall patterns of looking behaviours without inferential statistics to compare between conditions. Here I use the word 'exploratory' in a descriptive sense. This experiment will almost act as a pilot study to inform future hypotheses.

Despite this, in this experiment I assume that we will see the signalling 'dance' by participants; that being they spend more time looking to others while listening and less so when speaking (in line with previous research, see Chapter 1). I suggest that this pattern will be less recognisable in the triad condition. Here, I do not hypothesise that there will be a clear pattern of looking and listening behaviours. This is due to the complexity of the conversation (that being multiple speakers and listeners (meaning more possible interactions and signalling opportunities)) and variability amongst participants. For example, when speaking, participants have two people to potentially avert their gaze from and when listening, the attention may now be split between the two (perhaps linking to social referencing). Additionally, I was interested to investigate whether participants' visual attention is modulated by 'normal' (control) or averted eye contact. First, I question whether removing eye-contact affects looks to people. In line with Freeth, Foulsham and Kingstone (2013), we may expect fewer looks to the experimenter in the averted condition. Second, I explore how speakership affects this.

Method

Participants

Participants are 17 students and staff from the University of Essex who took part in return for payment.

Stimuli

Participants took part in a live conversation with an experimenter and confederate after completing a number of unrelated tasks around campus whilst wearing the mobile eye tracker. Participants were under the impression that they were being asked about the usability of the mobile eye-tracker. The questions used referred to how they found wearing the device including: “How did you find walking around? Did you notice people looking at you? Did you encounter any problems?”. The qualitative responses to these questions were not analysed and are not relevant for the current study’s aims.

Design

This study had 4 within-subject conditions: dyad pair with eye-contact, dyad pair without eye-contact, group of 3 with eye-contact and group of 3 without eye-contact. All participants took part in all conditions; hence this is a fully-within design.

Apparatus

Participants wore the SMI mobile eye-tracker during the conversation. Data collected from the mobile eye-tracker gives a scene view (taken from a small camera at the front of the participant’s head) and a fixation point which is overlayed on the video image (this is collected via two infrared cameras). For further details on mobile eye tracking see Chapter 2. Audio was collected via the scene camera mobile eye-tracker. Conversations were also recorded on two standard video cameras.

Procedure

Participants wore the mobile eye-tracker as they walked around campus and completed tasks (for example, 'please walk to the lake' and 'can you find a sign which says X'). After this walk, they were invited back into the lab to discuss how they found the task. The responses participants gave are irrelevant to this study aims, with this type of question used to ensure the participant found the conversation as natural as possible. At the point that this data was collected, the participants had been wearing the eye-tracker for roughly 40 minutes, in an attempt to habituate them to the apparatus (Nasiopoulos et al., 2015). First, the experimenter and the participant spoke alone for roughly three minutes. After this, a confederate was invited into the room, which the participant had passed multiple times when walking around campus. This changed the dynamics of the conversation from a dyad pair to a group of three. We did not counterbalance these conditions due to the nature of the task involving a 'surprise' confederate. Additionally, we had 2 further manipulations of with and without eye-contact (counterbalanced and fully presented within the group size conditions). For roughly half of each testing session (monitored by the number of questions asked), participants were offered normal eye contact by the experimenter. This Control condition meant the experimenter behaved naturally, which involved meeting and breaking gaze. In the remainder of the session, the experimenter looked at their notes and did not make any eye contact (Averted condition). Averted eye contact in this sense implies that the lead researcher would stop making 'normal' eye-contact for a portion of the recording. The researcher would instead stare at a piece of paper in front of them. The confederate continued with normal eye-contact.

These conditions were fully counterbalanced. This meant there were four testing conditions. The seating arrangements remained the same for all participants with the researcher and confederate seated at opposite ends of the table. This seating arrangement was

purposeful to make the participant looking behaviour more apparent (i.e., large head movements) by having a maximum distance between the confederate and experimenter (see Figure 3.1).

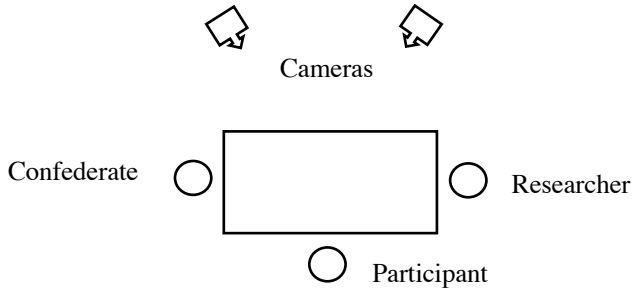

Figure 3.1. Image to show the schematic layout of the room.

For further details of the manipulations and their counterbalancing see Appendix 2. The total conversation for analysis for each participant was roughly 4 minutes in length (roughly 1 minute per condition).

Results.

Data preparation

Participants' initial testing sessions were first logged for key events which occurred throughout the recording. The key events included conversation starting, when eye contact was manipulated and at the stage the conversation progressed from a dyad pair to a group of three. Using Python, the clips were then cut down to the specific time frame of interest. Clips were then logged for participants' looking and speaking behaviour. The clips were manually coded using VideoCoder (1.2), a specifically designed programme for logging video clips, written by Tom Foulsham (University of Essex) in 2009. The manual coding was prepared by an additional researcher and a sample was compared to ensure reliability of coding. In line with the findings of face biases during MET coding in Chapter 2, clear rules of when a cursor

should or shouldn't be coded as 'on' a person were defined. Coding of the scene included gaze on and off the researcher and/or confederate as well as the start and stop times of speech for all interlocutors.

Exclusions

Prior to data analysis, four participants' data was removed due to visibly poor calibration, reported by the experimenter. Upon data analysis, it was apparent that a large amount of data was lost or had poor calibration despite extensive recalibration procedures throughout the testing sessions. For this reason, some participant conditions were removed from analysis. I did not choose to remove the data collected from the participant's full testing session if one condition was unobtainable. I also removed fixation and audio analysis from participants where the recorded testing session was much smaller than anticipated (i.e., 15 seconds). Out of a possible 52 test conditions (four conditions for 13 remaining participants), 43 were used in analysis.

Visual attention analysis

The log completed for each participant (which gave us speaking and gaze information), was processed in MATLAB to give one row of data per millisecond of analysis per person. This gave over three million rows of data to be analysed. From this extensive data set I was then able to calculate overall percentages for each participant, for each element of interest. Note that 100% here refers to 100% of the time that was manually coded within the condition.

Group Size Analysis

The results presented here first are averaged across both eye contact conditions (normal and averted).

Looking at group members

The amount of time participants spent looking at the experimenter and confederate was analysed. First, I calculated an average percentage time for each location (on experimenter, on confederate and elsewhere) for each participant. Overall, participants spent an average (SD) of 31.92% (38.36) of the time fixating the experimenter, 8.40% (22.95) of the time fixating a confederate (percentage calculated only from triad group) and 64.81% (40.40) of the time looking elsewhere. Note these percentages do not add to 100, given the percentage of time fixating the confederate is a percentage taken only from the triad condition. From Table 3.1 which shows the data split overall between group size, we can see there were fewer looks to a person in the triad condition (even when summing the looks to the confederate and experimenter together). However, as shown in Figure 3.2, there was large variability in the data.

Group Size	On Experimenter	On Confederate	Elsewhere
Dyad	36.06 (39.11)	NA	63.94 (39.11)
Triad	25.42 (37.19)	8.40 (22.95)	66.18 (40.40)

Table 3.1. The average (SD) percentage time participants spent fixating the group members during the conversation.

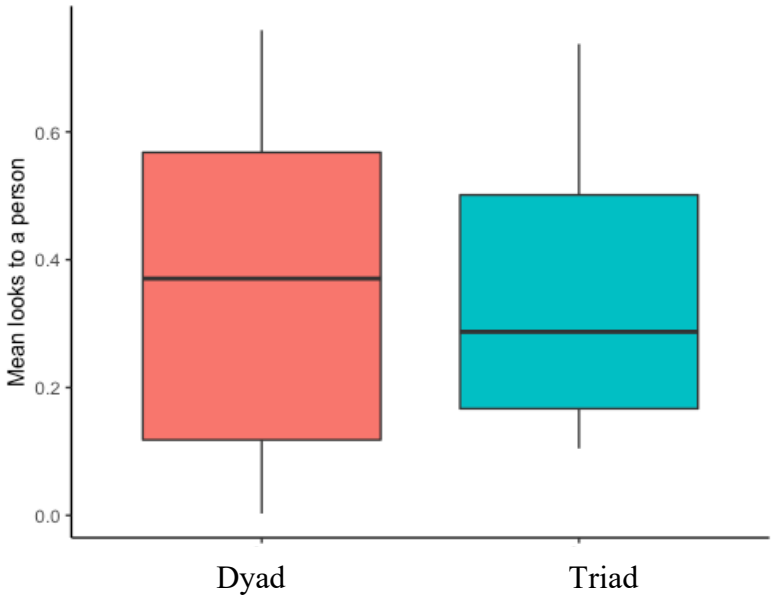

Figure 3.2. Boxplot to show the average time spent looking to a person (aggregated for experimenter and confederate in the triad condition). Boxes show the median and percentiles with whiskers showing the interquartile range.

Looks with speaking

Participants looking patterns were then compared with their spoken utterance throughout the conversation; specifically, this analysis examines the looking behaviour at the times of speaking and listening. 'Participant Talking' was calculated by taking the proportion of milliseconds the participant is currently speaking and 'Participant Listening' included data for times only whilst another person (confederate or experimenter) was speaking. Note that although the experimenter controlled the majority of the conversation, the confederate did speak occasionally. However as there were minimal looks to a confederate in general, I did not split this analysis further.

Overall, regardless of condition, participants spent 45.12% (39.43) of the time looking at others while they were listening to other speakers and 29.35% (36.07) of the time looking at others while they were speaking themselves. This supports the previous findings of increased gaze aversion while speaking (e.g., Ho, Foulsham & Kingstone, 2015). This is split by group size in Table 3.2. Here, on visual inspection of the data it seems average time

looking to a person is higher in the dyad condition, compared to triad. Figure 3.3 again shows the large variability that can be seen in the data.

Group Size	Participant Talking	Participant Listening
Dyad	30.95 (34.52)	46.26 (36.97)
Triad	26.85 (38.48)	43.44 (43.30)

Table 3.2. Shows average percentage (SD) looks to a person (either experimenter or confederate) while the participant is either speaking or listening, split by group size.

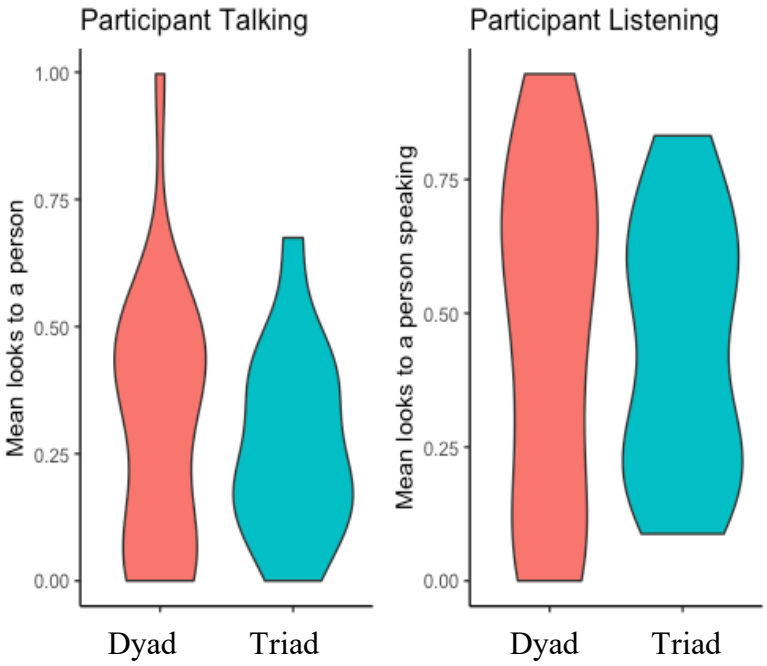

Figure 3.3. Demonstrates the mean percentage of looks to a person (experimenter or confederate) split by group size, whilst a participant is themselves talking (left), or listening (right). Note there are differences in the y axis scale.

Eye Contact Analysis

The results presented here are averaged across both group size conditions (dyad and triad) with 'person' referring to either the experimenter or the confederate.

Looking at group members

The amount of time participants spent looking at a person (experimenter or confederate) was first analysed. Table 5.1 shows the percentage time spent looking at the group members during the conversation split by eye contact condition.

Eye Contact	On a Person	Elsewhere
Normal	35.28 (47.73)	64.72 (41.17)
Averted	35.11 (45.81)	64.89 (39.70)

Table 5.1. Percentage time spent fixating group members in the two eye contact conditions.

On visual inspection of the data, there are minimal differences between the two conditions. As in previous analysis, data variability for average looks to a person was large (as shown in Figure 5.1).

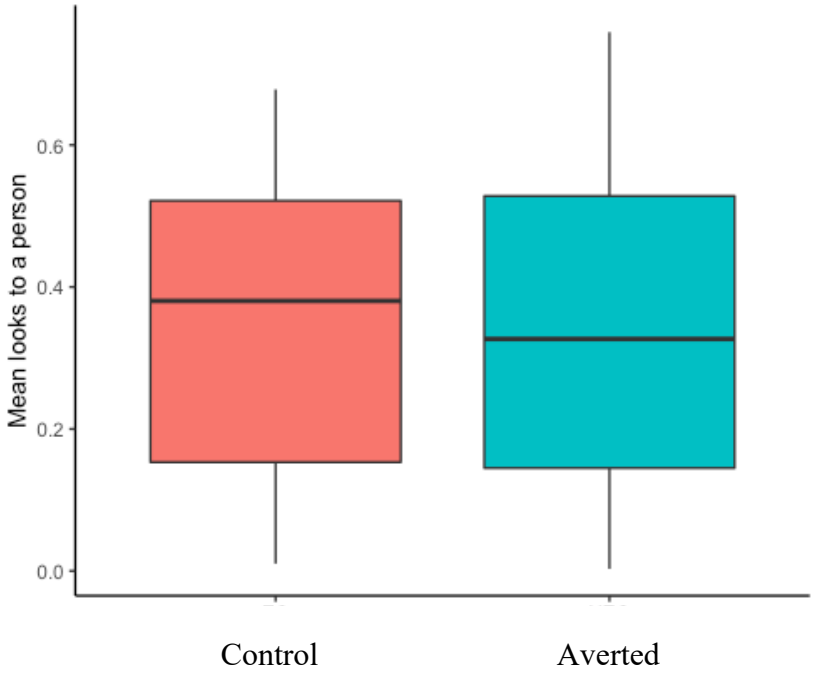

Figure 5.1. Demonstrates the variability of mean looks to person in the Control and Averted eye contact conditions. Boxes show the median and percentiles with whiskers showing the interquartile range Note here 'person' is the experimenter or the confederate.

Looks with speaking

As before, participants' looking patterns were then compared with their spoken utterance throughout the conversation; specifically, the analysis examines gaze at the times of the participant speaking and listening. For overall statistics see previous section. Table 5.2 demonstrates the percentage of time participants spent looking at others while speaking and listening, split by eye contact condition.

Eye Contact	Participant Talking	Participant Listening
Control	30.37 (39.16)	43.32 (39.15)
Averted	28.44 (33.29)	46.73 (39.68)

Table 5.2. Shows average percentage (SD) looks to a person (either experimenter or confederate) while the participant is either speaking or listening, split by eye contact condition.

The results demonstrate the same pattern (of increased gaze aversion while speaking compared to listening. However, when splitting this into the eye contact conditions, overall, we see slightly more social attention to people while listening in the averted condition. As well as slightly less social attention to people while talking in the averted condition. This suggests there may be an interaction. However, this is only a descriptive, numerical difference and Figure 5.2 showcases the large variability of this data.

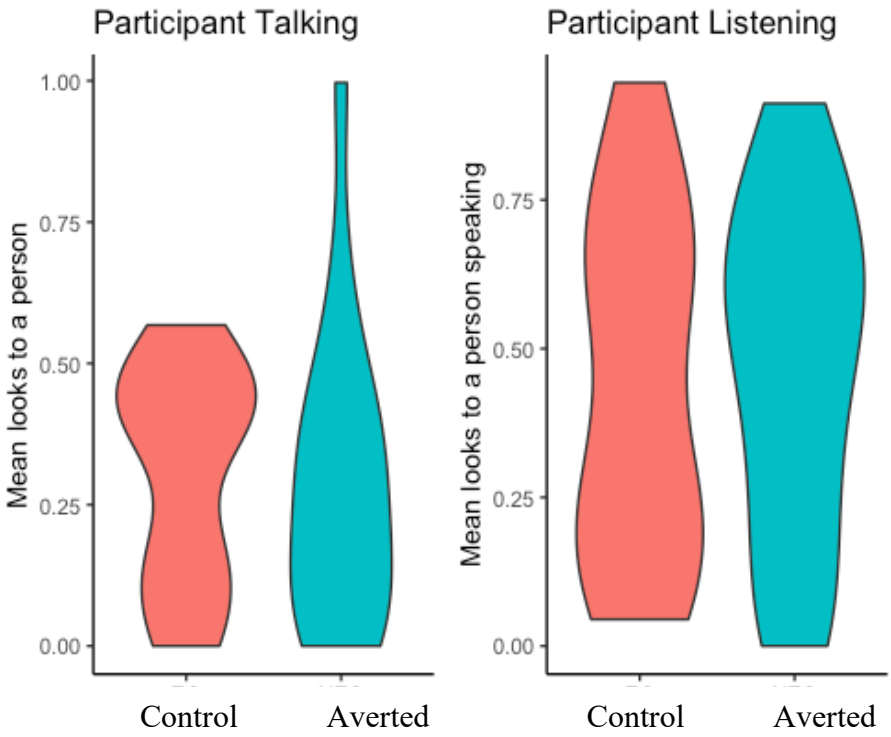

Figure 5.2. Illustrates the mean looks to a speaker when the participant themselves is talking and when the other group members are talking (participant listening), split by eye contact condition.

Discussion

This exploratory study investigated the patterns of looking behaviour when a participant forms a dyad and a triad and additionally whether there would be a different pattern of responses when an experimenter averts their eyes during a conversation. Overall findings suggest there were minimal differences in group size (and if anything, looks to a person decreased in larger groups) and minimal differences between eye-contact conditions, with large variability in the data. However, results support the notion of looking to a speaker while listening and averting the eyes while speaking.

Group size

Although no inferential statistics were run on this data, we see a trend of fewer looks to a person (overall) in the triad condition. This is irrelevant of whether the participant is speaking or listening. This contradicts results of Verteegal et al. (2001) who found an overall

increase in social attention in larger settings. However, this does support findings of Maran et al. (2020), who found that when engaging in a group there would be less gaze towards those participating than when in a dyad pair. There are a couple of suggestions as to why we might see this trend. First, perhaps in a larger group conversation it is harder to track who is speaking and where you, as a listener should be looking to signal you are paying attention. This could explain the results of the overall less looks to a speaker. However, this does not explain why in the triad condition there were less looks in general. Perhaps, given the experimental set up of the room, participants found it easier to look elsewhere (maybe straight ahead), when turn-taking conversation was occurring in the triad condition. A second suggestion refers to the term 'social loafing' (Maran et al., 2020) as described in Chapter 1, where there is less pressure to attend during group conversation. The results of this experiment would support this theory, given there was less visual attention to others when in a larger group setting. A third suggestion and given the large variability in the data, could be the participant differences in introversion. For example, it may be that larger groups made some participants shyer and hence less eye contact was made overall in the triad condition. It would be interesting to explore personality differences in future studies.

Eye contact

Despite minimal differences in eye contact conditions, upon inspecting the averages of the data, one interesting result is increased looks to a person with no eye contact when the participant is listening. As has been reproduced there is an overall increase in looks to a person when listening compared to speaking. However, it is interesting that when the experimenter does not make eye contact (and is speaking themselves) that there is increased social attention overall. Suggestions for this could include that the participant is using their eyes to make a stronger signal for reciprocal eye contact. The participant in a sense could be

looking more to the experimenter to cue or tempt eye contact from them. The small increase could also be from looks to a confederate instead of the experimenter (although there was only a small percentage of looks to confederates overall). A further point to note is that generally when a person is speaking (in this case the experimenter), we tend not to look at the listener anyway. Therefore, as we did not purposely manipulate eye contact in the Control condition, perhaps the experimenter would not have looked to a participant frequently whilst speaking anyway. Hence, the eye contact may have been averted without accounting for this. Alternatively, and a simpler explanation could be that this is a slightly 'odd' behaviour and the increase in social attention is because this is a slightly abnormal situation. This could have meant an overall unnatural visual attention distribution by the participant. In other words, perhaps the situation was strange, meaning the participant had decreased attention to the experimenter (regardless of eye contact). I did try to combat this limitation by habituating the participant to the eye-tracker (with participants wearing the apparatus for roughly 40 minutes before). However, as previously discussed, ecological validity is difficult to establish in social attention research.

Interestingly, when the participant was themselves speaking, there was less visual attention when the experimenter averted their gaze. If the opposite was found, it would suggest the participant was perhaps again trying to evoke eye contact or check the experimenter was listening to them. However, as the results in this experiment showed a slight decrease in social attention, perhaps the participant did not feel as engaged nor needed to check the experimenter was listening.

Overall, in terms of the theoretical implications of the findings presented in this exploratory study, we can suggest that there is an overall pattern of looking to an interlocutor less when speaking and more so when listening, regardless of manipulations (eye contact and group size). This supports previous research with similar live situations which allow for

reciprocal interaction (Ho, et al., 2015; Hessels et al., 2019) and demonstrates the strength of this effect even with these variations in the interaction. This further supports the notion that, in live interactions, the eyes are used as a signal and have a dual purpose (Gobel et al., 2015; Gregory et al., 2015).

Limitations

There are a number of key limitations within this exploratory experiment which could have affected results and should be reviewed.

First, a key limitation of this exploratory experiment is the high exclusion rate and the lack of useable data. For this reason, I chose to include this study within this thesis as 'exploratory'. Hence, here I only describe the results and hope to use this research to generate future hypotheses. The problems with the mobile eye-tracker are highlighted in Chapter 2, but authentically exist in the present experiment. Overall, the general data quality was poor. Despite cautious efforts to recalibrate and validate the eye-tracker (and accompanying video cameras as support), the data was poor, in particular when moving to a triad set up (perhaps due to increased participant movement). This also meant, despite the experiment designed in a way to allow within-subjects comparisons, this was not possible. For example, participant 1 may have had accurate data in the dyad condition and not in the triad condition. Rather than removing the participant as a whole, I decided to still include the data where possible. However, due to the large amount of missing data, direct within-subjects comparisons could not be made. For this reason, I am cautious to suggest any predictions were confirmed or refuted.

Furthermore, there was very large variability between and within participants. Although I had planned to conduct a time-series analysis, the large variability and poor data quality meant this wasn't possible. For this reason, a larger sample should be used in follow

up MET studies. This would not only help the power of the study, but also help with the high exclusion rates. In future studies, it may also be interesting to take an individual differences approach, to establish if other factors of the person (e.g., personality) modulate the variability of the results.

It should also be noted, as is found in my Experiments 1, 2 and 3, there is a strong bias to code cursors near faces as 'on' the face. This raises questions as to the fallibility of the coded data. Despite attempts made to ensure there were rules regarding the coding and multiple coders comparing logs, there still may be an overall bias to report the cursor as 'on' the people within the scene. What is even more thought-provoking is whether this face bias is moderated by eye contact. I would expect the bias may be more prevalent with eye contact versus no eye contact. This would be an interesting next step to explore coding subjectivity of MET data.

A further consideration, which was only reviewed upon watching the scene cameras of the participants, is the positioning of the confederate and participant. I chose to position the two at opposite ends of the table (Figure 3.1), with the idea that this would help me to establish a clear difference of where they are looking. Within a pilot test, this proved beneficial. However, when watching the scene videos back from participants, it became apparent that there are increased problems with calibration when participants move their head, particularly when they move their head fast. Furthermore, it could be argued that proximity may affect looks to the speaker and confederate. For example, social referencing may not occur (e.g., looks to a confederate when the experimenter is speaking) when the interlocutors are positioned further apart, as they were in this experiment. If you think about your own conversations, it may seem odd to turn you head to the other side of the room when one person is talking to you, just to check their response. However, when they are located closer together, perhaps more social referencing could occur.

I would also choose to manipulate eye contact using sunglasses (see Experiment 7 and 8 for the impeding effects of occluding the eyes with sunglasses). This would remove any unwanted head movements or other factors (i.e., an experimenter acting differently) and should help to isolate the effect of the eyes as a signalling cue. As stated, this descriptive experiment was created with the view to run a confirmatory follow up study with clear hypotheses. In terms of future eye contact hypotheses based on the current results I would predict that eye contact may increase social visual attention. It would also be interesting to explore how speech of the participant is affected by eye contact

It should also be noted that additional MET studies were planned to explore gaze in conversation (see Appendix 5 for an example). However, due to the pandemic, these additional studies could unfortunately not be carried out. Experiment 4 was hence an exploratory experiment, with confirmatory research to follow. In confirmatory studies, a clear hypothesis and power testing would have been generated from this present research. My future hypotheses generated from Experiment 4 would be that when moving to larger groups, attention to others in the interaction is decreased overall. If I was to run the confirmatory research, I would explore this with the participant making half of dyad pair or a group of four. This would allow a clearer comparison of the visual attention allocation. I would also test a larger number of participants to allow for omissions.

Conclusions

The exploratory Experiment 4 investigated how group size and eye contact affects looking to interlocutors. Overall evidence suggests there is a trend of looking more to people while listening and less so whilst speaking, supporting prior research. In terms of group size, it seems there was less time looking overall to people when forming a triad than a dyad.

There were minimal differences in eye contact. However, this data should be accepted with caution given extensive variability and methodology issues.

Chapter Summary

Here, I presented an exploratory experiment which explores eye movement in a face-to-face live setting. In doing so, I was able to investigate how multiple people interacting affects looking behaviours. In other words, both the experimenter, confederate and the participant were able to signal with their eyes, something which cannot be explored using videos as stimuli. Given the results of Experiment 4, I have demonstrated how mobile eye tracking (which enables 'real-world' data) is more difficult to collect, analyse and make generalisations from. For this reason, using mobile eye tracking data is not necessary for questions within Chapters 4 and 5, where I instead use videos shown to third parties. This enables me to collect data with greater accuracy. Additionally, within Experiment 5 (Chapter 4), I demonstrate the close link between live and third-party visual attention, which justifies the use of videos in subsequent chapters. However, after exploring the criticisms of stimuli often used in social attention research within this chapter, the videos in the subsequent experiments were created with careful consideration with aspects including using larger, naturally formed groups and free flowing dialogue.

Chapter 4: Which audiovisual cues guide visual attention during conversation?

Experiment 5 within this chapter is published in a special issue of Visual Cognition, with Tom Foulsham as second author.

In everyday group conversations, we must decide whom to pay attention to, and when. This process of dynamic social attention is important for goals both perceptual and social. The experiments within Chapter 4 specifically explore how audio and visual elements of conversation act as cues to guide attention in group social settings. Moving away from general visual attention in social situations, this chapter focusses on deciphering which audiovisual cues evoke gaze shifts and fixations.

Two experiments (5 and 6) investigate the manipulation of such cues in third-party observation to explore gaze to targets and conversation following. Experiment 5 explores how similarly conversation is visually followed in live and third-party observers. In the third-party observers I manipulate the availability of both visual and auditory elements of the clips, and hence the accompanying signalling behaviours. Experiment 6 explores the strength of this sound-image association in Experiment 5's freeze-frame condition when the targets' location is spatially manipulated. Using a combination of interest area and time-series analysis I paint a clearer picture as to how attention is directed when observing group interactions.

Experiment 5 – Your turn to speak? Audiovisual social attention in the lab and in the wild

Experiment 5 gives a further informed account of using audio and visual information in gaze and is inspired by the described research depicting the disparities in importance of audio and visual modalities in guiding conversation following. The experiment uses a natural group interaction to a) explore how third-party visual attention to this group interaction is modulated by manipulations of modalities present and b) to explore to what extent the third-party visual behaviour maps to live eye movements at the time the stimuli was formulated.

Introduction to Experiment 5

In a world which is full of complex social scenes, the ability for us to selectively attend to targets of the most importance is a rapid, fluid, and sophisticated process. A considerable amount of research has helped us determine the systems involved in selectively attending to social elements in static images, but less is understood about the processes involved in observing dynamic situations, in particular within complex social settings. A relevant observation when thinking about our own everyday interactions, and one that is easily replicated in the lab, is that we should attend to someone who is speaking. We look to a speaker not only to aid language comprehension but also as a signal to show we are listening. From childhood, we are often told "look at me when I am speaking to you" and having our eyes on the teacher indicates that we are listening.

To converse efficiently, an exquisitely attuned, adaptive and coordinated system is required to process the dynamic information present (Penn, 2000). In the current study, we investigate such dynamic processing during participation in live group conversations and compare this with third-party observations of those conversations. In particular, we examine gaze behaviour as a way to investigate the perception of social cues which allow people to

follow a conversation. For more information regarding audio and visual cues, see prior literature.

Third-party versus live interaction

The present experiment uses a combination of pre-recorded conversations shown to participants and a live situation to help us to understand how visual attention is distributed during social interactions. We include this comparison as the extent to which findings from studies with pictures and video can be generalised to real life is debated (see Risko et al., 2012, for a comprehensive review). Studies which have explored to what degree a lab scenario reflects a 'real' situation have uncovered distinct differences in social behaviour. For example, Hayward et al. (2017) demonstrate differences in social attention engagement between a real-world task and a more typical lab-based social cueing task. This is further echoed in work by Foulsham and Kingstone (2017) who demonstrated a fairly poor relationship between real-world gaze behaviour and fixations on static images of the same environment. Risko et al. (2012) explain how we should exercise caution when drawing conclusions solely from findings using static stimuli in a controlled experiment. Risko et al. (2016), advocate 'Breaking the Fourth Wall' within social attention research, to enable a method which is more representative of a real interaction. Their paper argues that social attention research has often failed to recognise that in a real-world scenario, the person or agent within the scene can interact with the participant, while a pre-recorded video or image cannot. This interactive element will clearly have dramatic effects on participant behaviour.

A previously discussed instructive example is given by Laidlaw et al. (2011) explored to what extent participants look at a person if they are in the room or are presented on a video camera. Participants were more inclined to look at a person on a video than in real life, demonstrating the effect of real social presence on visual attention. Foulsham, Walker and Kingstone (2011) found that when comparing the eye movements of people walking through

a campus with participants watching those clips at a later stage, although there were similarities, the live presence of other people did affect gaze. For example, pedestrians who were close to the observer were looked at more by observers watching the event on video than by people in the real world. Social norms may play a role in these discrepancies, but the differences can also be explained in terms of real versus implied social presence and the previously discussed dual purpose of the eyes. Any differences, which could be explained by the signalling of the eyes in a live situation, will be investigated here by comparing between a live interaction and responses to pre-recorded video.

Experiment 5 research questions

The present experiment investigates the signalling cues utilized in visual attentional shifts during turn-taking conversation, whilst offering a unique method to compare a live scenario with people watching a recording. To our knowledge, this experiment is unique in allowing the comparison of live and third-party group gaze behaviours. We have three main research questions.

Does third-party viewing reflect live gaze behaviour?

First, we explore how visual attention differs in the lab and in a natural conversation using methods which will allow for a comparison which has not previously been available. We do this by recording video stimuli during a naturalistic interaction. In line with previous research (such as by Freeth, Foulsham & Kingstone, 2013 and Laidlaw et al., 2011) we may find less looks to speakers in the live situation than in the video observations due to social avoidance. Alternatively, if the reason we look to a speaker is to signal to others that we are listening, it could be argued that looks to speakers may increase in a live situation compared to a video; something which would be redundant in third-party observers. If we see similar results in both situations this would indicate that we do not look at a speaker just to signal, instead perhaps this is a habit or aids comprehension in some way. Comparing the

interlocutors' visual attention with that of third-party observers watching the same conversation on video will also help establish the ecological validity of understanding social attention via pre-recorded videos, which will add confidence that our additional research questions are relevant for real interactions. To assess this, we will examine the degree of looking to current speakers, other targets and elsewhere in both a live interaction and when watching a video, and we will additionally evaluate the 'agreement' between looking behaviour in each setting over time.

How do audiovisual cues affect conversation following?

Second, we test the signalling cues which attract our visual attention to a speaker during videos of group interactions. Previous studies have examined the impact of removing the sound (Foulsham & Sanderson, 2013). We will additionally manipulate the visual information available to the participant by freeze-framing the image or transitioning to a blank screen while the audio continues. These conditions will be compared to a Control clip where both audio and visual information is available. The research by Hirvenkari et al. (2013) suggests that participants will continue to look at the image of the person speaking, even in the freeze-frame condition where no information from their movements is available. This might occur because observers have built an association between an individual voice and that person's face, although whether this has benefits for comprehension remains to be established. Participants might even be linking the voice to a spatial location on the screen. Therefore, this experiment will also explore the association between audio and spatial location with a blank screen condition inspired by past work on 'looking at nothing' (Richardson et al., 2009). This line of research explores how participants have a tendency to look to the blank space where stimuli were previously presented when later hearing information relating to those stimuli. Altmann (2004) proposes that this is due to a 'spatial index' which is part of the memory representation of the object.

As people look to social elements of a scene, we hypothesise that the bias to fixate the person speaking may be reduced but still present in the freeze-frame condition. Participants may not show a pattern of following the speaker in either of the visual manipulation conditions, as no additional visual information can be gained. However, if participants adopt a 'looking at nothing' approach a pattern of conversation following may remain. This will help us to uncover how the auditory component of conversation allows for conversation following and whether observation of the targets (moving, frozen or not at all) is crucial for this. We will investigate the effect of the sound being removed, the image freezing or the image being completely removed on looks to people, their features (eyes and mouth), and in particular the time spent looking at the current speaker.

When are speakers looked at?

Third, this experiment aims to investigate the precise timing of looks to a speaker. There are some inconsistencies in previously discussed findings regarding this time course, with most studies showing that participants' gaze anticipates changes in the conversation turn but others observing that there is a lag between the utterance beginning and fixation on the speaker. Often, the research which demonstrates an anticipatory effect involves stimuli depicting just two individuals. This might facilitate conversation following, in that participants can easily distinguish who will be the next speaker. The evidence for third-party anticipation in larger groups is limited and less consistent. However, we expect that participants will continue to shift gaze to the speaker prior to the utterance beginning. If the anticipatory effect is equal in all conditions, this would suggest that participants use visual and auditory elements equally to guide their attention to speakers. However, if participants rely on auditory cues, we expect the anticipatory effect to be diminished in the Silent condition. However, if participants rely on visual elements (e.g., head and eye movements of the depicted people) to induce an early gaze shift, we would expect there to be no

anticipatory effect in the freeze-frame and blank condition. To explore this, we will analyse at which point participants make a fixation to current speakers upon that utterance beginning.

In addition to our three research questions, we aim to test these aspects of social attention in a more complex environment than the one in which they have previously been studied. The research conducted to date is often scripted with dyad pairs, and fewer studies have considered larger group interactions, even though these are common in everyday life. The present experiment uses naturally formed groups of 6 individuals, all of whom were members of sports teams at the University of Essex, seated around a table to enable a fluid discussion. Adding additional people to the group and the use of free-flowing conversation comes with increased complications for analysis, but also increased visual attention decisions which need to be made by the observer, providing us with rich multimodal data. With multiple targets and multiple turn-taking transitions, we might expect more variation in observer gaze and perhaps less attuned timing patterns of conversation following. In a dyad pair paradigm, often used in third-party eye tracking studies which include audiovisual modulations, the decision for the interlocutors involved is only whether to direct or avert one's gaze. In a larger group, one must decide who to look at, and when (for example distributing attention between the speaker and people who are listening). Although we still expect a speaker will dominate fixations in both the live and third-party participants, with a larger group gaze may be more distributed than previously reported. In addition, using a larger group setting will add supporting or contradictory evidence of predicting utterance starts within a complex social environment.

Materials and methods

The aims and analysis of this experiment were pre-registered (see https://osf.io/m2dp5/).

Participants

We here analyse data from small groups interacting in real life, as well as from participants later watching video recordings of these groups. The individuals recorded in the real interaction are hereafter referred to as the "targets" and the third-party participants referred to as "participants".

The targets were drawn from 4 groups of 6 individuals comprised of various sports teams at the University of Essex. There were 2 groups of males and 2 groups of females. An initial request to take part was sent to the Presidents of the sports clubs. A full description of this interaction and the target recordings is provided in the next section. For analysis of behaviour from the group interaction, and for creating stimuli, we relied on data from one half of the table. There were therefore 12 targets in these clips. This sample size was predetermined based on the required stimuli for the second part of the experiment.

The third-party, eye tracked participants were 40 volunteers (7 male and 33 female), with a mean (standard deviation) age of 20.9 (2.9) years old. This sample size was pre-registered and with this within-subjects design gives excellent power for effects such as the one of sound on gaze in Foulsham & Sanderson (2013; $dz = 2.4$). All participants were undergraduate students from the University of Essex, recruited for course credit. All participants had normal or corrected-to-normal vision and gave their informed consent before taking part.

Target clip preparation

The video clips shown to third-party participants depicted 6 individuals having a discussion while sitting around a table, with only 3 individuals (one side of the table) in view in each clip. This is shown schematically in Figure 4.1.

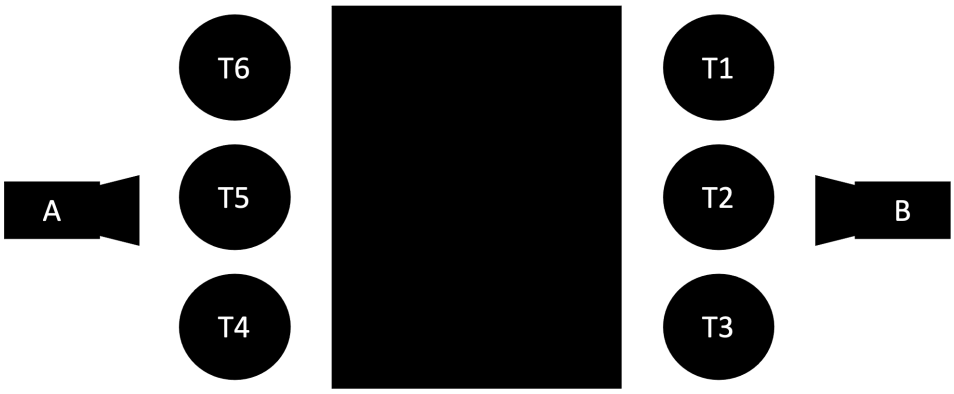

Figure 4.1. Schematic view of target individuals (T1-T6) and video camera set up during stimuli creation.

The clips were derived from a 1 hour recording with two static video cameras (with microphone) placed discretely, which are a permanent feature of the Observation Laboratory at the University of Essex. Each camera was adjusted so that only the view of one side of the table was present within the recording. The view from camera A was used to create the clips for the eye tracking experiment. The view from both cameras A and B were used to code behaviour of the targets in the live interaction. The discussion took place in a well-lit room. The targets were given several questions, in a randomised order, which they were to discuss as a group. The questions given to the targets were questions or topics designed to enable natural conversing from all team members. Examples include: 'find out who has moved house the most', 'what are you most grateful for?' and 'what is your most embarrassing moment?'. Two experimental clips were selected from each continuous recording and featured moments where all visible targets spoke at least once. These clips were selected to ensure that Targets 1, 2 and 3 were the predominant speakers, with minimal involvement from the targets on the other side of the table. Additionally, a 'familiarity clip' was prepared for each target group. This clip also featured the visible targets all speaking at least once.

These clips were included to ensure that the third-party participant was familiar with each of the targets' voices. The familiarity clips were not used in further analyses. Hence participants were shown a total of 3 clips per target group (12 clips total). Clips varied in length from 36-54 seconds.

The audiovisual information in the experimental clips was manipulated to produce a Control condition and 3 alternative conditions: Silent, Freeze Frame and Blank. For these conditions, a critical time when the manipulation began was chosen for each clip. This was roughly mid-way through the clip but at a point when all 3 target faces were clearly visible (i.e., not covered by hands), and given the range of durations the time of the manipulation was unpredictable for the participants. For the same reason, we did not count an exact number of turn exchanges prior to the manipulation and instead chose a point when there was fluid conversation. In the Silent condition the audio was removed at this point while the video continued. In the Freeze Frame condition, the visual image was frozen and in the Blank condition, all visual stimuli were removed, and a plain white screen was presented. In both the Freeze Frame and Blank conditions, the audio continued (see Figure 4.2 for a visualisation). The videos were re-encoded to a frame resolution of 1024x768 pixels and displayed at a rate of 25 frames per second.

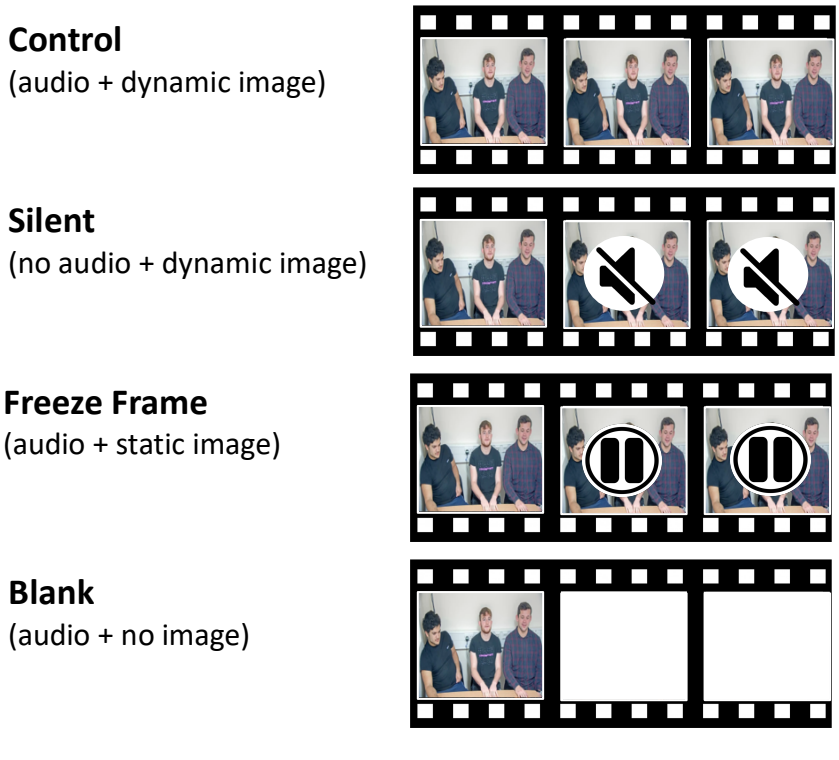

Figure 4.2. A visualisation of the 4 video conditions (Control, Silent Freeze Frame and Blank) shown to third-party participants.

Apparatus

An EyeLink 1000 eye-tracker and ExperimentBuilder software were used to present the stimuli and record eye movements. Monocular eye position was recorded by the Eyelink 1000 system by tracking the pupil and the corneal reflection at 1000 Hz. A nine-point calibration and validation procedure ensured mean gaze-position errors of less than 0.5 degrees. Saccades and fixations were defined according to Eyelink's acceleration and velocity thresholds.

The video clips were presented on a 19inch colour monitor. During all conditions, sound was played through headphones which the participants were required to wear throughout the experiment. The participants' head movements were restricted using a chinrest which kept the viewing distant constant at 60cm. At this distance the video frame subtended approximately 38° × 22° of visual angle.

Participant procedure

The participants read and completed consent forms and were asked to confirm that they had normal or corrected to normal vision before beginning the experiment. Participants then took part in a 9-point calibration. After the participant's right eye had been successfully calibrated and validated, the experiment began.

There were 4 blocks of trials, with each block consisting of clips from one of the target groups in one of the 4 conditions. Participants watched 3 clips per block, with the single familiarity clip always preceding the (2) experimental clips. Participants only saw each clip once, but clips were counterbalanced across the conditions, such that over the whole experiment each clip appeared in each condition equally often. This ensures that any

peculiarities of the particular clip or the following question could not explain differences between conditions. Condition blocks were presented in a randomised order.

Participants were simply instructed to watch the scene and not given any further instructions about how to view the scene. Participants were informed that they would be asked a question after each clip, based on what they had seen. After watching each clip, participants were given a simple comprehension question based on the conversation the targets were engaging in. The questions were in the style of "Which person said X", with participants responding to the questions by pressing '1', '2', or '3' on the keypad to indicate one of the three targets. Questions were based on each particular clip, but again these were randomly ordered and counterbalanced across conditions as described above. The questions were piloted for difficulty before beginning the experiment and only used as an attentional check to ensure participants were paying attention to the clips. After each clip there was a drift check which ensured accurate tracking throughout. The overall testing session (of eye-movement collection) lasted approximately 9 minutes.

Results

This experiment is unique in that it offers a comparison between the third-party viewing of a conversation (which we would expect from a typical eye tracking study) with gaze at the time of recording the stimuli in a live situation. We begin by making this comparison as a manipulation check, to first assess how visual attention in the live interaction compares with third-party viewing in the lab. Then we progress to the effects of audio and visual modalities on conversation following. Data and scripts for our analysis can be found at https://osf.io/m2dp5/.

Does third-party viewing reflect live gaze behaviour?

Behaviour in the live interaction

During the live interaction, all six targets in each of the groups were filmed during the interactions. Third-party viewers, when watching the manipulated clips, only saw one side of the table (Targets 1, 2 and 3 as seen in Figure 4.1). However, the fact that the other side of the table was also filmed during data collection enables us to analyse the live viewing behaviour of the 3 targets not present in the stimuli (Targets 4, 5 and 6). This gives us live visual attention data for 3 observer targets per group (a total of 12). When collecting the video prior to beginning the conversation, targets were asked to systematically look at each person in the room. This gave us a 'calibration' to which the researcher could refer when making decisions regarding coding where the target was looking.

Clips of the live behaviour from the other side of the table were trimmed to the exact time of the 8 experimental clips used in our main experiment. By choosing these exact moments of conversation we can make a comparison between the gaze of people sitting in the room with the targets and the gaze of our eye tracked participants who later watched the videos. Of course, there are differences between these two sources of data because we only have 3 people sitting opposite each target in real life, and we rely on coding their gaze from video, compared to a much larger group of eye tracked participants. In this section we therefore focus on describing similarities and differences rather than null hypothesis significance testing.

We logged the time at which each utterance began and ended alongside where each target was looking (for targets 4,5 and 6) at each point in time. To accurately log when the utterances began and ended, we used the auditory signal with the visual signal to assist in identifying the speaking target. We found we could reliably determine when the targets (T4, T5 and T6) were looking at T1, T2 or T3 (located opposite). Gaze to these locations was

clearly visible and accompanied by head movements. To code the looking behaviour, we used VideoCoder (1.2), a custom software tool designed for accurately time-stamping events in video. Gaze locations were then manually categorised according to which target was being fixated and whether that target was currently speaking. This log was prepared by the author and one other naive researcher, with high interrater reliability of the coding (98% agreement of sample compared).

Time series for gaze and speaking were analysed in MATLAB. We removed the small percentage of time in which the monitored targets were speaking themselves. Overall, averaged across the 12 individuals and 8 clips which we coded, targets spent 59.04% of the time looking at a target on the other side of the table who was currently speaking, 34.88% of the time on a non-speaking target and 6.09% of the time looking elsewhere (such as at the table or their own hands).

Comparing live interaction with third-party viewing

We began by comparing gaze in the live situation with that from the Control condition, in which third party observers watched videos of the same targets without any audiovisual manipulation. Here we are comparing manual coding of looking behaviours (live interaction), with eye tracked data, collected in the lab, from the Control condition (third-party viewing). Table 4.1 compares the proportion of time that the interacting targets spent looking at the speaker with the same percentages from third-party video watching.

Gaze location	Live interaction - all targets combined	Third party viewing (Control condition)
% on a speaking target	59.04	51.06
% on a non-speaking target	34.88	47.91
% elsewhere	6.09	1.04

Table 4.1. Average percentage time spent in each of the gaze locations for the live behaviour and the third-party viewing of the Control condition.

In general, the percentages are quite similar, with a majority of time spent looking at the person speaking in both the live and lab situation. We make this comparison cautiously, given the differences in in sample size and data collection, and refrain from a statistical comparison. It may be that the bias towards the speaker is reduced in third-party participants watching the interaction on video, who look at the non-speaking targets more than in the live situation. A potential explanation for this is that in a live situation, we tend to look to a person speaking. We do this to signal that we are listening (e.g., Freeth et al., 2013). When watching a recording, this signalling is not possible, which perhaps meant that participants felt more able to visually explore the other targets.

Comparing timing of looks

For a more in-depth assessment of the similarities in looking behaviour between the live interaction and participants watching the videos, we compared the time series of who was being looked at in each case. Figure 4.3 illustrates one example, in which we compare the 8 participants (P1-8) who saw the first clip in the Control condition, with the 3 targets (T4, T5 and T6) who were present in the room. Gaze time series are displayed for each observer (right panels), while we also plot the proportion of observers looking at each target at each point in time (left panels). It is clear that observers are highly consistent, both within a group and between the two environments. For example, most observers shift gaze from Target 3 to Target 1 about a third of the way through the clip. Target 2 receives less attention, with looks to this person peaking at the end of the clip.

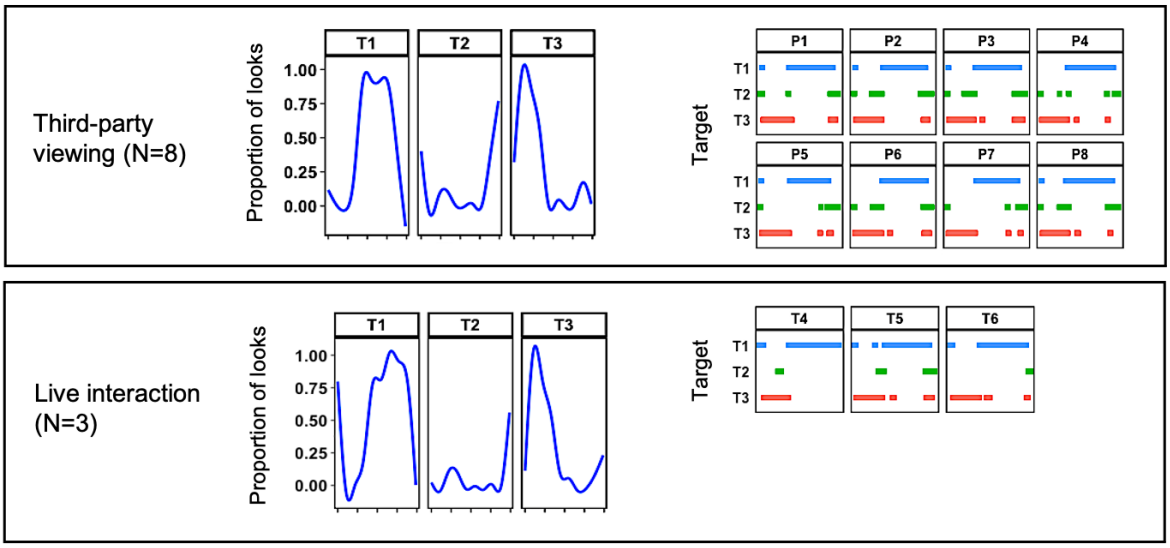

Figure 4.3. Time series representing the gaze location of each eye tracked participant (P1-P8, third-party participants) and each interacting target (T4-T6, live interaction) as they looked at the targets of interest (T1-T3). Line charts on the left show the proportion of observers gazing at each location (data smoothed over time). Coloured bars on the right show the target being looked at by each observer. In each case, time is on the x-axis (clip duration =39000 msec). Within this example there is an average of 88% agreement between live and third party (average $\varkappa$=.76 with all pairings p<.001).

Measuring agreement

To test the strength of gaze 'agreement' (the extent to which those in the live and third-party condition were looking at the same target at the same time), we calculated the amount of time when the gazed-at location was the same in each pair of observers, comparing each live observer to each video-watcher in the lab. We excluded times when observers were looking elsewhere (this included blinks or data loss in the eye tracked data and times when the real-life observers were looking down or at people on their own side of the table).

When analysing the time series, across all combinations of pairs in each clip, an average of 60.10% of looks were to the same target (T1-T3) at the same time, which is greater than chance (33.3% of all visual attention if this was shared equally between targets T1, T2 and T3). Additionally, a Cohens kappa analysis was run to determine the strength of agreement in gaze. The kappa statistic provides a proportion of agreement over and above

chance when comparing nominal data. In general, there was substantial to strong agreement, with all combinations providing a kappa coefficient which was significantly different from chance. Across the 8 experimental clips, there was a mean (SD) kappa of .79 (0.1), with all combinations reaching at least a kappa of .59. For a more detailed look at these results see Appendix 3.

Together, these analyses demonstrate that people view the conversation in a similar way whether they are in a real situation taking part in the conversation themselves or different participants watching the conversation on video at a later stage. This is reassuring as it suggests that gaze behaviour to the videos will be a good proxy for complex social attention in a face-to-face situation.

Eye tracking experiment

We have established that there are similarities between visual attention in conversation following when watching videos (Control condition) and in the wild (real interaction between targets). Observers tend to look at the speaker and show consistent patterns over time related to the conversation. We now test how these patterns are affected by manipulations of audio and visual cues, as well as examining the timing of gaze in the detail afforded by a controlled eye tracking experiment. Due to the nature of the live interaction, we cannot compare how audiovisual cues affect visual attention during live setting. The remainder of the analysis is therefore for third-party participants only. As described in 'Target Clip Preparation', we had 4 audiovisual conditions: Control, Silent, Freeze Frame and Blank. If the tendency to look at the speaker at a particular time is dependent on auditory information and/or visual cues such as gestures, then we should expect this to be reduced in the Silent and Freeze Frame conditions, respectively.

Outliers and exclusions

All participants scored over 50% on the comprehension questions, with a minimum accuracy of 58% and an overall mean of 88% correct, so no participants were excluded on this basis. However, 3 participants were excluded due to a failure to calibrate and validate the eye-tracker within satisfactory parameters, resulting in frequent missing data. For this reason, 37 participants were included in the main analysis. For data investigating the general oculomotor measures see Appendix 3.

How do audiovisual cues affect conversation following?

The presented analysis considers gaze behaviour only after the critical time point in each clip (i.e., from the point at which the clip was silenced, frozen or blanked), and the equivalent time in the Control condition. This critical time point varied slightly across the 8 experimental clips but was identical for all conditions within a particular clip. Due to the differences in the critical time period, we report durations as a percentage of the total length of this period.

Fixations on targets

To explore how the condition affected how much participants looked at the three targets, a region of interest (ROI) was defined around each target individual that was present. The ROIs subtended approximately 10° × 11.5° of visual angle, see Figure 4.4, however this varied according to the physical size of the target within the scene. For comparison, in the blank condition, the same ROIs were used even though the image had at this point disappeared.

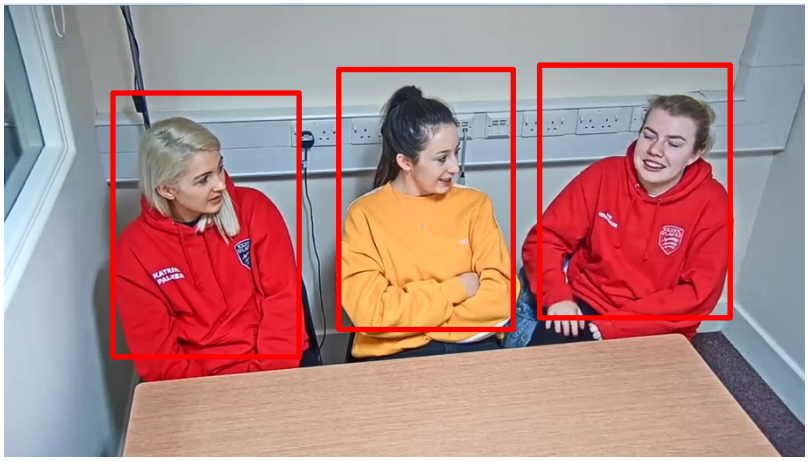

Figure 4.4. An example video frame, with ROIs selecting each of the three targets.

The number of fixations on these ROIs was then analysed. Pooling these ROIs together, we found that, regardless of condition, participants spent the majority of time looking at the targets rather than elsewhere on screen (see Table 4.2). For analysis per target member see Appendix 3.

	Control		*Silent*		*FF*		*Blank*	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
% fixations on targets	98.19	2.2	97.14	4.8	93.68	8.8	72.10	22.4

Table 4.2. Overall percentage of fixations on targets, post clip manipulation (average taken from 37 participants).

A repeated measures ANOVA established that there was a significant difference between the percentage of fixations on targets in the different conditions, F (3,108) =43.43, p<.001, η_p^2 =0.55.

Pairwise comparisons with a Bonferroni correction (SPSS adjusted p values), revealed that the percentage of fixations on target ROIs in the Blank condition was significantly lower than the other conditions (all t (36) > 6.67, p <.001, dz >1.10). This is perhaps not surprising given that the targets were no longer visible in the Blank condition. There was also a significant difference between the Control and the Freeze Frame conditions (t (36) =3.03, p

=.027 *dz*=0.63), with slightly fewer fixations on the targets when the video was paused. Although the Freeze Frame and Blank conditions elicited fewer fixations on targets, participants still looked at these regions on 94% and 72% of their fixations, respectively. Hence, despite there being no new visual information available at this point (since the video had paused or disappeared), participants continued to fixate the location where the targets had been for the majority of the time. It could be argued that the targets take up a large proportion of the screen although the targets do not consume an area of the screen approaching these high percentages (roughly 42%). For data analysis of fixation to target faces see Appendix 3.

Fixations on targets' eyes and mouth

Previous studies have found differences in fixations on targets' eye and mouth regions when sound information is removed, but these have been small or inconsistent (Foulsham & Sanderson, 2013; Vo et al., 2012). For this reason, in our next analysis, moving ROIs were created for the 8 experimental clips using Data Viewer (SR research). An ROI was drawn around each of the targets' eye and mouth area and its position throughout the recording was adjusted by slowly playing the clip back with 'mouse record' (an inbuilt function in Data Viewer). For comparison, the location of these interest areas remained at the same location in the Freeze Frame and Blank conditions (i.e., at the location of the eyes and mouth when the video paused or was removed). Fixations were then analysed to determine whether they were inside this area.

Figure 4.5 shows the percentage of all fixations on the targets' eyes, mouth and elsewhere averaged across participants. It is important to note that 'elsewhere' includes the rest of the target and background areas.

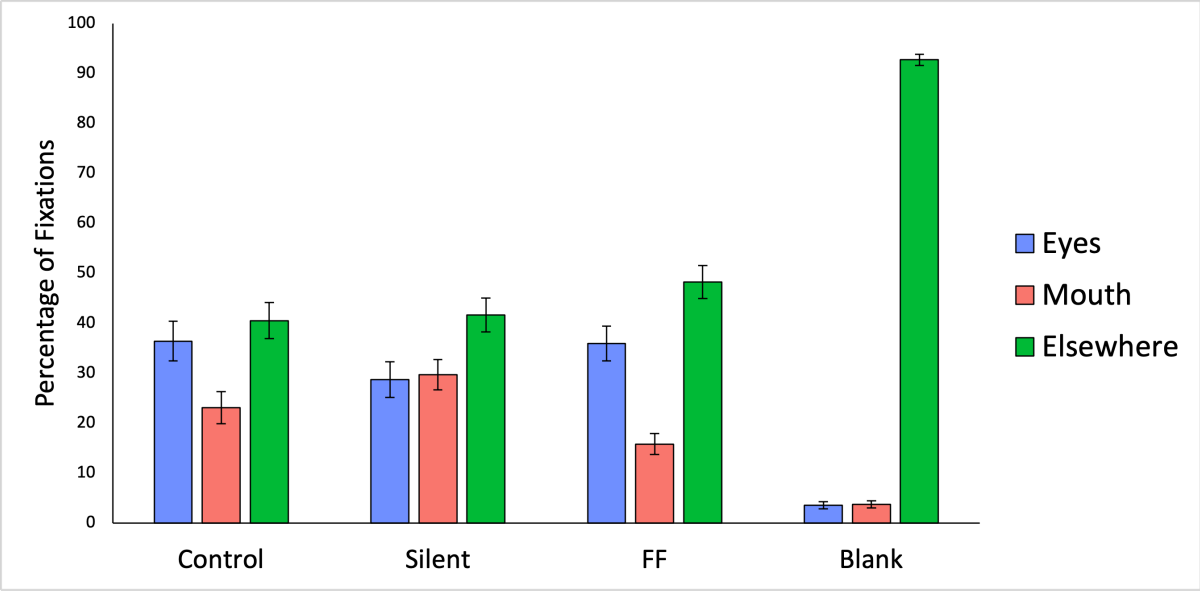

Figure 4.5. The overall percentage of fixations on the targets' eyes, mouth and 'elsewhere' (sum per condition = 100%), averaged across the participants. Error bars show standard error.

A repeated measures ANOVA established that there was a significant difference between conditions when analysing looks to the eyes, F (3,108) = 40.94, p<.001, η_p^2 =0.53. Post-hoc analysis with the Bonferroni correction (SPSS adjusted p values), revealed there were no significant differences between the Control, Silent and Freeze Frame conditions (all t (36) <2.50, p>.10, dz < 1.33). Hence, the looks to eyes did not significantly decrease with the sound muted, or with the image stilled. The only condition with a different pattern was the Blank condition which was significantly different from all other conditions (all t (36)>7.26 p<.001, dz >1.24).

When analysing looks to the mouth, a repeated measures ANOVA established that there was a significant difference between conditions, F (3,108) = 30.88, p<.001, η_p^2 =0.46. With a Bonferroni correction (SPSS adjusted p values), there were significant differences between the Blank condition and all other conditions (all t (36)>5.97, p<.001, dz >1.11). There were no significant differences when comparing the Control condition with the Silent and Freeze Frame conditions (both t (36) < 2.61 p>.79, dz < 1.18). Looking at the average

percentages of looks to the mouth, we can see that removing the sound did slightly increase looks to the mouth (and reduce those to the eyes), which is similar to the pattern reported by Foulsham and Sanderson (2013). However, as in that study, the difference was not significant.

Fixations on speaking targets

Next, we used the previously described record of who was speaking to examine how condition affected the tendency to look at the speaker. Figure 4.6 shows the total proportion of fixations which landed on a target who was currently speaking, grouped by condition.

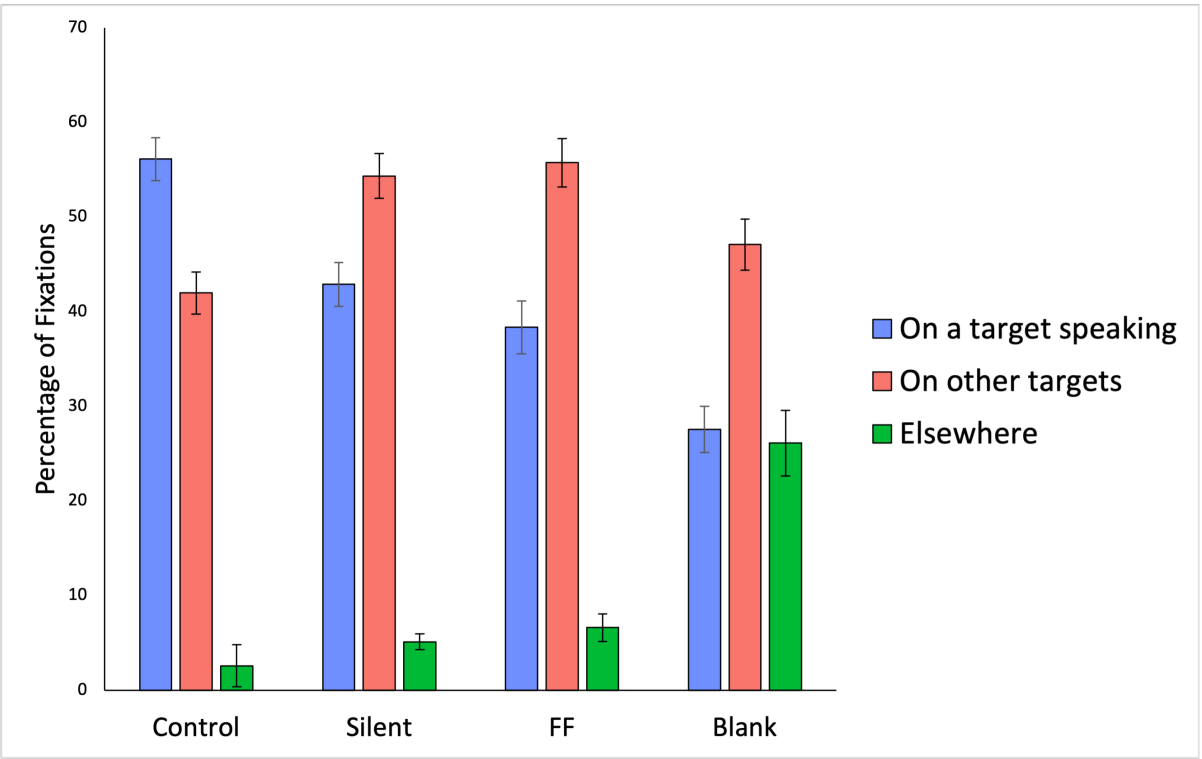

Figure 4.6. Percentage of fixations on target speakers for each condition (note 'other targets' refers to the two other non-speaking targets grouped together). Error bars show standard error.

A 3x4 fully within subjects ANOVA was carried out for percentage of fixations on the (3) regions of interest and in each of the (4) conditions. There was no significant main effect of condition, however there was a main effect of fixation location F (2,72) =338.18, p<.00,1 η_p^2 =0.90. More importantly, there was an interaction between condition and fixation

location F (6,216) =18.31, p<.001, η_p^2 =0.34. This emerged since the distribution of fixations to locations in the Control condition was different to the other conditions. In this condition, most of the fixations were on the person currently speaking, with fewer on one of the (non-speaking) targets (even though this comprised two different targets). The Control condition is most similar to the live viewing behaviour, which we discussed earlier and has a similar ratio of looks to speakers and non-speakers. In contrast, fixations on the speaker were reduced in the Silent and Freeze Frame conditions and reduced further in the Blank condition.

When are speakers looked at?

Timing of fixations on speaking targets

We then analysed the point in time at which participants made a fixation on a speaker, in order to understand whether conversation following varied with the cues provided in the 4 conditions. The start times of each utterance from each target in each clip were used to create 10ms bins ranging from 1000ms before speech beginning to 1000ms after speech beginning. We then compared these bins to the fixation data and coded bins as to whether they contained a fixation on the target speaking, a fixation elsewhere or no fixation. The result was an estimate of the probability of looking at a speaker, time-locked to the beginning of their speech. Figure 4.7 plots this estimate averaged across all clips and utterances and split by condition.

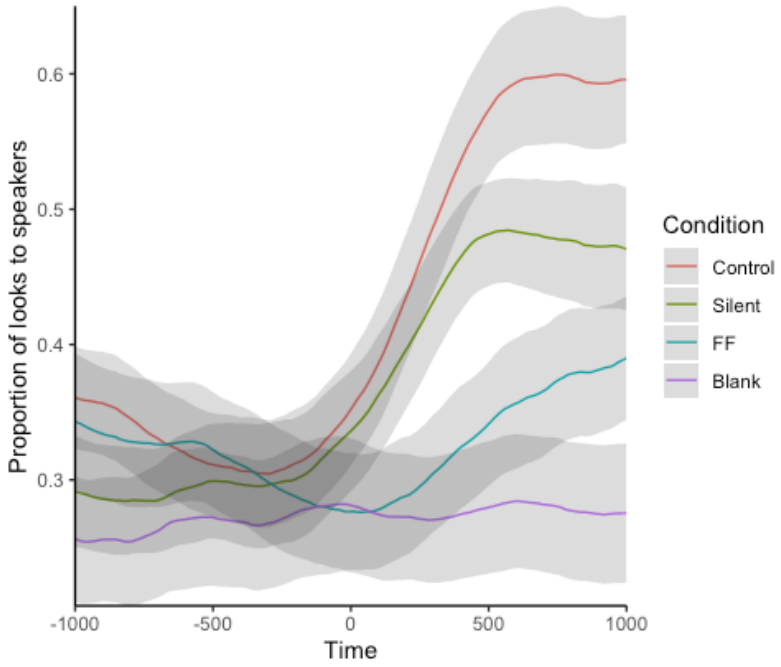

Figure 4.7. Probability of fixation being on the speaker, relative to when they started speaking. Lines show the smoothed, average proportion of fixations at this time on the speaker, in 10ms bins (with 95% CI). A time of 0 indicates the time at which a speaker began speaking. FF: Freeze Frame condition.

As previously discussed, from Figure 4.7, we can see there are fewer looks to speakers in the modified conditions in comparison to the Control condition. There are, however, clear increases at the time of speech onset in both the Silent and the Freeze Frame condition, which indicates that participants are still shifting their visual attention to that target. The slope of this increase is similar in the Silent and Control conditions. Although the slope of the Freeze Frame condition is more gradual, it still rises to a peak indicating increased attraction to a speaking target.

To quantify the time at which the probability of fixation diverges between conditions, we ran a cluster-based permutation analysis between each pair of conditions. This is a non-parametric method for comparing timecourses while controlling for multiple comparisons (Dink and Ferguson, 2015). The results indicated that most of the conditions diverged significantly close to the moment the utterances began (between -50ms and 60ms relative to

when the speaker began talking). Interestingly, as can be seen in Figure 4.7, the visually dynamic conditions (Control and Silent) diverge from each other much later (with a significant difference after 450ms). The visually static conditions (FF and Blank) also diverge later (after 570ms). This implies significant increases in proportion of looks to speakers at the time of utterance when dynamic visual information is present, and a slower effect of the divergence between the combinations of the visual and auditory information.

Discussion of Experiment 5

The present experiment provided an innovative way to evaluate viewing behaviours during a conversation in a live situation and when observers watched a recording. In line with previous research, this experiment found that most visual attention was directed towards the targets, with little to no attention to background areas. There was also a tendency to look to the person currently speaking. This was apparent for both real interactants and third-party observers. There were a number of interesting findings relating to manipulation of signalling cues and the timing pattern of looks to targets, which we here discuss with reference to our three research questions.

Does third-party viewing reflect live gaze behaviour?

First, we investigated visual attention during a real group conversation and how it relates to third-party viewing. There is existing observational research investigating gaze in face-to-face conversation (see Hessels, 2020 for an extensive review), and previous studies have used eye tracking while people watch pre-recorded conversations (e.g., Foulsham et al., 2010; Tice & Henetz, 2011; Foulsham & Sanderson, 2013). However, to our knowledge, this is the first time that gaze in a real multi-party conversation has been explicitly compared to fixations from people watching videos of the same interaction. Our findings demonstrated

that in a live situation, people sitting opposite their interlocutors show a similar pattern of gaze behaviour to participants who were watching the conversation at a later stage. For instance, the tendency to look at the person who is currently speaking was similar, comprising 59% and 51% of the time, for live and third-party viewing, respectively. The similarities in gaze distribution in the two settings indicate that the sole purpose of looking to a speaker isn't to signal that we are listening, as we see this effect in the third-party video setting which doesn't require any social signalling. Therefore, looking to the speaker must benefit conversation following for another reason or may be a habitual behaviour. It is not surprising that these percentages are somewhat less than previous studies (such as Argyle and Ingham (1972), as the present experiment included a complex group interaction, rather than a dyad, with more targets for participants to distribute their attention between.

The temporal pattern of looks to different potential targets was also strikingly similar between the real interaction and the video watching condition. Different individuals tended to look at the same target at the same time, and this was also true when we compared eye tracked observers with the people who were actually in the room with these targets. There was a high level of agreement in all cases. It therefore seems that visual attention to conversation in live interlocutors and third-party observers shows a similar pattern. Future studies could also explore to what extent there is a lag when comparing the timecourse of looks to speakers in real interactions and when watching video.

Although there are good reasons to think that social gaze operates differently in face-to-face situations (Risko et al., 2012; Risko et al., 2016), it may be that the pattern of conversation following observed here is rather unaffected by actual social presence. This would indicate that investigations where third-party participants observe conversations on video provide a good test-bed for realistic social attention. On the other hand, we did observe some differences between settings that could be pursued in future research. On average, third

party observers spent more time looking at targets who were not speaking, than those in the face-to-face situation. This could be explained by social norms, which are only present with social presence when interacting face to face. For example, in a live group conversation it might be considered odd to look at someone who is not speaking, when there currently is another member of the group speaking and when collectively the group attention is on the speaker. However, if you are watching a video recording of the conversation, this social rule is not present, and participants may have felt freer to explore the reactions of other targets to the speaker. In larger groups, this might be particularly prevalent so that listeners can check group agreement or to monitor other's reactions to the conversation. This relates to previous early work such as by Ellsworth, Carlsmith and Henson (1972), who demonstrated the discomfort which arises when being stared at. Additionally, this dovetails with research by Laidlaw et al. (2011), where it was demonstrated how the presence of a confederate (versus the same confederate on videotape), increased visual avoidance. Critically, in the video condition in both Laidlaw et al. (2011) and our experiment, participants could not see or interact with the targets, and so any "signalling" function of gaze was absent (Risko et al., 2016).

Interestingly, we saw striking similarities in eye movement behaviour in the two settings even though in the live situation, the targets were acquaintances and in the third-party situation, the targets were strangers. In future studies it would be interesting to explore the effect of familiarity on both the live eye movements and in the third-party participants (for example using observers from the same sports team).

Overall, we analysed visual attention within a larger group setting, with multiple targets and multiple turn-taking transitions. Despite some differences in attention to non-speaking targets, those who were taking part in the conversation and those who watched the same conversation at a later stage still show a bias to the speaker. This suggests that the

distribution of gaze is similar in both settings and furthermore suggests social gaze in a complex environment is similar to that of studies which use dyad pairs. This provides further evidence for the ecological validity of understanding social attention via pre-recorded videos.

How do audiovisual cues affect conversation following?

An advantage of using video recorded stimuli, of course, is that they can be controlled and manipulated in a way that real interactions cannot. In the present experiment we manipulated conversation video clips in order to test the role of audio information (by removing the sound) and dynamic and static visual information (by freezing the image or completely blanking the display while the sound continued). After these manipulations occurred, participants spent 98%, 97% and 94% of the time looking at the targets in the Control, Silent and Freeze Frame conditions, respectively. Percentages this high are not particularly surprising considering that the targets were the main focus of the scene, and that previous evidence indicates that observers are biased to look at people. In the Control and Silent condition, the targets were also the only moving objects within the scene. However, the result in the Freeze Frame condition is particularly interesting as no new visual information could be gained from continuing to fixate the targets. In short, the tendency to look at the targets in these clips was not due to either the audio information or the dynamic visual information which would allow integration of speech and vision. This is explored further in Experiment 6.

In the Blank condition, the percentage of fixations on where the targets once were dropped to 72%. Although the targets were the main focus of the scene, the ROIs around the targets took up less than half of the screen area. Hence it could be argued that even in the Blank condition participants did continue to fixate the location of the targets once the image had been removed. This could relate to previous research on 'looking at nothing'. For

example, Richardson and Spivey (2000) demonstrated that when participants are shown a blank screen and asked to recall information, there is a tendency for them to look to the space in which the recalled item was once located. The authors argue that this is linked to memory of spatial location and the cognitive-perceptual system's ability to attach a spatial tag to a semantic location. A second study which supports this notion of visually "following" or attending to an object which is not visually present, is that by Spivey et al. (2000). That study found that when participants are passively listening to audio of a story which includes directionality, (such as a train moving from right to left), saccades follow the same pattern, in that the eye movements cluster along the same axis, even when the eyes are closed. The current experiment provides some evidence for spatial indexing in that participants may have associated voices and the mental model of the conversation with locations on the screen. As a result, blank regions where targets had previously been located remained salient to look at. In the Freeze-Frame condition, participants may also have created an association between the voice of the target and their spatial position on screen. Future research could investigate whether this association and related eye movements might facilitate the participants' understanding of the scene.

We then explored whether our clip manipulations affected looks to the current speaker. Foulsham and Sanderson's (2013) study, which used a similar methodology, found that participants looked most to a person currently speaking (both in their Control and Sound Off condition). Arguably, the results in their Sound Off condition could be due to the participant attempting to lip read, or to the fact that movement attracts our attention. We questioned whether we too would find this effect in our Silent condition and whether manipulating the visual information would affect looks to speakers. We found that participants fixated to speaking targets for 56% of fixations in the Control condition and 43% in the Silent condition and comparably to Foulsham and Sanderson (2013) removing

the sound did slightly increase looks to the mouth. In comparison, Vo et al. (2012) report that removing the audio decreased looks to the mouth region. This may be explained in terms of the task. In the present experiment, there were functional benefits of looking at the mouth (to provide a correct answer on the attention check question), whereas Vo et al. (2012) asked the observer to rate likeability which may not have required conversation understanding. Interestingly, in the current experiment, when the video was paused, participants continued to look at the eyes with a frequency similar to the normal Control condition, while their looks to the mouth decreased.

Although there was no additional information gained, participants fixated speaking targets (or the space where they once resided), for 38% and 28% of fixations for the Freeze Frame and Blank condition, respectively. The low percentages aren't surprising in the Blank condition. However, the Silent and Freeze Frame condition do show some evidence to suggest that people continue to track a speaker without dynamic or audio information. The differences between the Control and Freeze Frame condition could be explained in that without any dynamic visual information (i.e., observing the targets' mouth moving), it is difficult to determine the current speaker from audio alone. However, even in the Freeze-Frame condition, there were more fixations on the current speaker than we would expect if attention was allocated equally to all the targets. This suggests that people may attempt to look at a speaker upon hearing their voice to gain more information, as in spatial indexing during the 'looking at nothing' phenomena. Perhaps attaching the voice to a static image of the speaker provides richer understanding of the conversation and helps us to explain why we see a gaze shift towards a speaker.

Further work should attempt to understand to what extent and why do we continue to associate the audio of the target's voices with their spatial location and what benefit, if any, there is to this behaviour.

When are speakers looked at?

Foulsham et al. (2010) reported that gaze tends to precede or predict a change in speaker, such that observers look at conversants slightly before they start to speak. Similarly, Tice and Henetz (2011) found results which demonstrate anticipatory looks to a speaker (taking into account the 200msec to plan and execute an eye-movement). Gaze also tends to precede speech in real face-to-face conversations, at least in dyads, where speakers look at a listener at the end of their "turn" in order to signal a change in speaker. However, it is less clear cut as to whether this happens in group conversations and whether these anticipations rely on particular audiovisual cues. For example, Holler and Kendrick (2015), report that when interacting in a group of 3, interlocutors are able to anticipate speakership, yet Foulsham and Sanderson (2013), who use a similar methodology to the present experiment, report no anticipatory effect. That study investigated the role of speech sound on gaze to speakers, by showing clips of group conversations to participants. The participants fixated the speaker quickly after they began speaking, but there was no evidence of a preceding shift. We expanded upon this to investigate whether an anticipatory effect would be present and whether this was affected by the presence of auditory cues (such as the content of the conversation) or visual signals (such as gestures or expressions before a new speech turn). In the present experiment, consistent with Foulsham and Sanderson (2013), no anticipatory effect was found in the Control condition, with fixations on speaking targets occurring on average 450-500ms after the start of an utterance.

There could be a number of reasons for this finding when using 'natural' conversation as stimuli. First, the type of clip used in the current experiment was quite different from Tice and Henetz (2011) despite both including a group conversation. In Tice and Henetz (2011) the stimuli used are from a Hollywood movie, 'Mean Girls', which comprises of a split screen dialogue. Such clips are designed in a way to guide our visual

attention to the most critical areas of the scene, through the use of camera angles and cinematic effects. This is likely to make it very clear who is speaking and when. For example, when a new character speaks, the screen splits further, directing your attention to the new element within the scene. For this reason, conversation following (and in this case anticipation of speaker) is facilitated by the editing (for an in-depth computational model of gaze trajectory in staged conversation see Boccignone, et al. 2020). In the present experiment, the conversations were unscripted, un-edited, and more complex, reflecting a real-life chat amongst friends. Targets often interrupted or spoke over each other and hence, perhaps in a real-world situation with multiple sources of sound, it is more difficult to predict the next speaker.

Despite there being no evidence of an anticipatory shift, participants do move quickly to fixate a target who begins speaking (see Figure 4.7). There is evidence for this in the Control condition, and to a lesser extent in the Silent and Freeze Frame conditions. Compared to the Control condition, there is a similar rise in probability of looks over the time course in the Silent and Freeze Frame conditions. This adds further evidence that participants do follow the conversation without the use of the full set of audiovisual cues. We therefore suggest, in audiovisual recordings when one modality is redundant, participants rely on the signalling cues available within the other modality to follow conversation.

Conclusions

The present experiment offered a chance to investigate audiovisual cues and make a comparison between live and third-party viewing behaviours, during a large group conversation. We demonstrate that live viewing behaviour during interactive conversation is similar for those taking part in the conversation and separate observers at a later stage,

highlighting the propensity to follow speakers during conversation in both situations. We further emphasised the ability for participants to exploit cues in both the spoken conversation and the movements of targets to follow turn-taking in conversation, even when one modality is removed. Removing the audio replicated the results of Foulsham and Sanderson (2013), with participants able to follow only visual cues to guide their attention. In addition, there is evidence for a tendency to look at the speaker, even when no additional visual information is gained (when the scene is frozen). The results provide insight into using audio cues to direct our attention, as well as how and why we observe dynamic and complex group engagement scenes, in a setting of more naturalistic composition. Overall, this experiment provides us with rich information about how visual attention is directed within multifaceted large group conversation with both manipulated videos and interactions during a live conversation.

Experiment 6 – Does the target's spatial location affect visual attention to conversation?

The following experiment expands on the previous conclusions of Experiment 5, exploring the finding of how there is continued visual attention to a target when the visual scene is static. This is an interesting finding, considering no additional information can be gained. Experiment 6 explores this further with additional freeze-frame manipulations to question if I can replicate this finding (looking to targets when they are frozen) and whether the expected spatial location of the target affects fixations to the speaker.

Introduction to Experiment 6

In the previous Experiment 5, we established that participants continue to look at a video of people talking, even when the image has been stilled. Although there was a slight decrease in looks to targets, the percentage of looks was still far greater than chance. In other words, this suggests participants could use the audio alone to follow the conversation on screen. This finding echoes work such as by Hirvenkari et al. (2013).

Questions arose as to the robustness of this effect which the present experiment explores. First, I question if this effect can be replicated reliably, and whether a different experiment paradigm specifically designed to test this phenomenon would reproduce similar results. Second, in order to tease apart why this happens, I question whether participants would also continue to fixate these targets even when the spatial location of the targets has been manipulated.

For this reason, the stimuli for this experiment were designed to optimally test these posed questions. The present stimuli involved participants watching clips of a group of targets which were presented in boxes, allowing only the targets' head and torso to be present. The remainder of the scene was a black background. I chose to manipulate the spatial location of these targets once frozen, in order to see if conversation following continued with

moving the expected location of the targets. Hence attempting to investigate if participants look to frozen targets due to association during conversation following (i.e., fixating a current speaker's frozen image as they associate their voice with their frozen image). I do this by presenting clips either in a typically expected horizontal orientation or placing boxes of targets in a stacked presentation (see Figure 4.8 in Methods). More specifically, as I am interested in whether spatial movement affects how participants look to a target (after the clip is freeze-framed), I will compare clips where target boxes have and have not moved, in the same spatial orientation.

Predictions are that participants will continue to look to the targets on screen for the majority of the time. If participants continue to look at the target who is currently speaking, (regardless of spatial changes) this will determine that looking at the face of the speaker (despite not gaining any additional information) seems to provide some sort of benefit. For example, it could be that looking at the target face is due to association of face and voice, or recognition. Equally, it could be that looking to the faces of speakers aids comprehension of the conversation.

If I find participants continue to look to where targets 'should' or once resided (after a spatial change), this could relate to the 'looking at nothing literature'. This phenomenon is described as the ability for the visual system to construct and store internal memory representations with the act of looking at nothing facilitating the ability to retrieve those representations (Ferreira, Apel & Henderson, 2008). For example, Richardson and Spivey (2000) demonstrate participants make saccades to empty regions where semantic information has previously been held and this is further supported in similar memory tests.

Together, with Experiment 5, the results will help us to understand why participants look at static images of group conversation when audio continues.

Methods

Participants

Participants were 36 students (25 female) from the University of Essex recruited for course credit. Participants had a mean (SD) age of 23.58 (6.64) years All participants had normal or corrected-to-normal vision and gave their informed consent before taking part. Participants were not permitted to take part if they had already taken part in a previous experiment (Experiment 5) using the same type of video clips.

Stimuli

Target clip preparation

The video clips used were the same clips in Experiment 5, with additional manipulations. See Experiment 5 for information on how the initial clips were recorded.

For this experiment, two experimental clips were selected from each continuous recording (six clips total) and featured moments where all visible targets spoke at least once. Clips were 35 seconds in length.

These six clips were manipulated six different ways. Initially the clips were manipulated so that a box was cropped around the three visible targets faces and upper body regions. The boxes were all edited to be the same size. These boxes were either kept in horizontal order as they would normally appear on the screen [1] or were manipulated so that the two targets on the side were now placed on top of each other (stacked) [2].

A second manipulation was that the clips either played continuously within these two conditions, or the clip freeze-framed [3]/[4], whereby the sound continued and the clip was stilled. The final clip manipulation included changing the spatial area of the speaking targets, from either control to stacked [5] or stacked to control [6] at 20 seconds into the clip when the clip stilled. This final experimental manipulation makes a total of six clip manipulations for each of the six experimental clips. Hence, 36 different stimuli were prepared. Each

participant saw one example of each clip manipulation (six experimental clips). Figure 4.8 demonstrates the six experimental clip conditions in a schematic format.

[1] Horizontal Control (the clip plays through).

[2] Stacked Control (the clip plays through, but the expected spatial location has changed).

[3] Horizontal Freeze-Frame (the clip pauses at 20000 msec).

[4] Stacked Freeze-Frame (the clip pauses at 20000 msec).

[5] Horizontal to Stacked (at 20000 msec the location of the targets changes and freezes).

[6] Stacked to Horizontal (at 20000 msec the location of the targets changes and freezes).

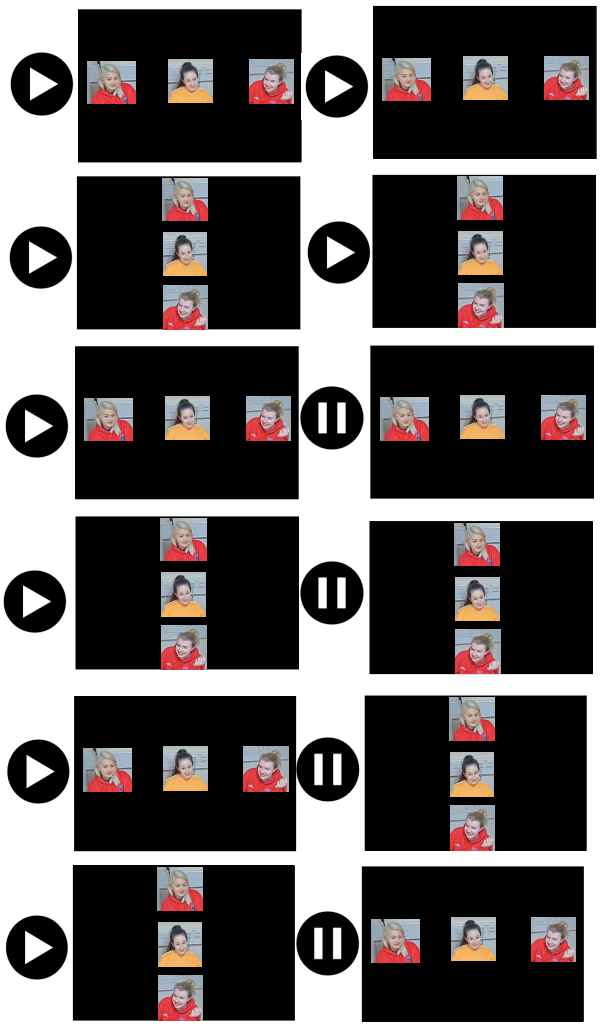

Figure 4.8. Figure to show the six clip conditions schematically.

The participants were also shown a non-manipulated clip, referred to as a 'Familiarity Clip' which was a clip for the participants to familiarise themselves with the targets' voices and appearance. This clip was always shown prior to the experimental clips. This clip was not used in analysis.

Attention questions were also prepared for each clip. This was a question about the content of the target's conversation. This was included to ensure the participants were paying attention but was not used in analysis.

Design

Participants were only shown each clip (1-6) once. Hence, they only saw one of the six conditions per clip. Due to this, six versions (A-F) were prepared, and participants were randomly assigned one of these versions (with equal numbers for each version). The clips were shown to the participants in a randomised order, but always with the Familiarity Clip preceding the relevant experimental clip.

Apparatus

The apparatus used to build and record eye movement was the same as used in Experiment 5.

Procedure

The participants read and completed consent forms and were asked to confirm that they had normal or corrected-to-normal vision before beginning the experiment. Participants then took part in a 9-point calibration. After the participant's right eye had been successfully calibrated and validated, the experiment began.

Participants were shown a Familiarity Clip of each of the three groups followed by experimental clips. Therefore, participants were shown a total of six experimental and three Familiarity clips. Participants were simply instructed to watch the scene and not given any further instructions about how to view the scene. Participants were informed that they would be asked a question after each clip, based on what they had seen. After watching each clip, participants were given a simple comprehension question based on the conversation the targets were engaging in. These were the same questions as prepared in Experiment 5. The questions were only used as an attentional check to ensure participants were paying attention

to the clips. After each clip there was a drift check which ensured accurate tracking throughout. The overall testing session (of eye-movement collection) lasted approximately six minutes.

Outliers

During data collection, one participant failed to complete the experiment and was excluded. For this reason, an additional participant was tested. Preliminary checks of the data found a lack of fixation data for more than 50% of recording time in one participant. This participant was therefore removed from the analysis. The results section therefore is for 35 participants.

Results

General oculomotor measures

First, I was interested in the general viewing behaviour of the participants. I included this analysis to confirm participants' general eye movements were not abnormal given the slightly obscure nature of the stimuli. Table 4.3 demonstrates the general oculomotor measures averaged for participants across clips.

	Control		Freeze-Frame only		Freeze-Frame with change in spatial location	
	Horizontal (1)	Stacked (2)	Horizontal (3)	Stacked (4)	Horizontal to Stacked (5)	Stacked to Horizontal (6)
Fixation Count	73.69 (15.78)	79.77 (19.56)	82.54 (18.58)	84.91 (15.81)	84.09 (17.81)	85.06 (17.92)
Saccade Count	72.83 (15.75)	79.03 (19.51)	81.71 (18.61)	84.26 (15.74)	83.40 (17.71)	84.34 (18.13)
Fixation Duration	428.70 (115.66)	391.50 (149.81)	362.63 (126.38)	347.16 (96.50)	391.90 (102.80)	359.01 (102.55)

Table 4.3. General oculomotor measures averaged (SD) across participants, per clip.

Target interest areas

Next, I was interested in whether participants continued to look at the targets in all six conditions and how the looking behaviour to targets changed with these clip manipulations. Interest areas were drawn using DataViewer's built in IA analysis around the 3 target boxes. In both of the Control and Freeze Frame conditions (horizontal and stacked where no spatial moving occurred), these remained constant. In the conditions where target's location spatially changed on the screen, this interest area moved at 20000msec. As I am only interested in the time after which the manipulation occurred, the remainder of the analysis will only look at visual attention post 20000msec.

First, the number of fixations that each participant made past the critical point (post-clip manipulation) were quantified. The number of those fixations which were on an interest area (any target) were then calculated for each participant. Overall, an average of 89.68% of fixations past this critical time point were on targets, regardless of condition. This is split by condition in Table 4.4.

	Control		Freeze-Frame only		Freeze-Frame with change in spatial location	
	Horizontal (1)	Stacked (2)	Horizontal (3)	Stacked (4)	Horizontal to Stacked (5)	Stacked to Horizontal (6)
Average % of fixations ON targets	93.34 (16.22)	95.62 (11.82)	87.14 (13.32)	93.00 (8.78)	87.60 (11.01)	83.65 (14.63)

Table 4.4. The average (SD) percentage of fixations on targets for each condition, averaged across participants.

When first looking at these averages, it seems the participants looked slightly less to the targets when spatial location changed (5 and 6). However, participants are still looking at the targets the majority of the time, despite gaining no additional visual information. A 1 x 6 repeated measures ANOVA established there was a significant effect of condition, F (5,170) =5.57, p<.001. Pairwise comparisons with a Bonferroni correction established there were only significant differences between the Control Stacked (2) condition with the two spatial manipulation conditions (5,6) and the Freeze-Frame Stacked (4) condition with the Freeze-Frame Stacked to Horizontal condition (6). The comparisons which were of particular interest to us were those where the clip content is the same (e.g., both either horizontal or stacked at 20000msec), but one of the clips included a prior spatial change. For example, comparing condition 3 with condition 6 and condition 4 with condition 5 (see Figure 4.8 for details). In both of these comparisons, when the clip includes a spatial change, participants do look less to the targets, although not significantly so. Hence, the spatial change appears to have a small but non-significant effect on looks to targets. This implies participants continue to look at targets equally, regardless of whether a spatial change has occurred.

Out of curiosity, I also analysed whether any of the 'elsewhere' fixations (fixations not on a target) landed on where the targets previously resided prior to the spatial change. For this analysis, I looked at only conditions where there was a spatial change (5 and 6). I did not

consider looks to Target 2 (middle target in each clip), as this target did not move, with only the peripheral Targets (1 and 3) changing location. On average less than 3% of fixations landed where the targets previously were, demonstrating no effect of looking back to prior semantic locations (see Figure 4.9 for an example). Despite this, when looking at Figure 4.9 we do see more evidence of looking to where targets once resided than to the outer corners of the screen.

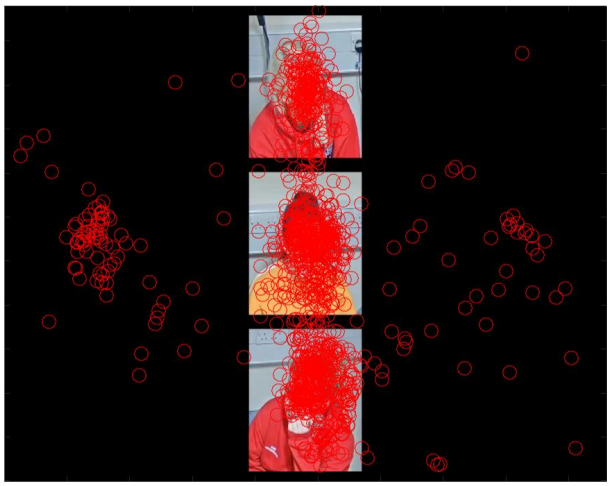

Figure 4.9. Demonstrates an example of fixations after clip manipulation. Red circles indicate fixation locations. Note that this is for demonstration purposes and features all fixations in this condition for multiple clips.

Looks to target speakers

Finally, I wanted to know whether there would be a difference in looking to a target who is currently speaking. I was interested to know whether participants continue to associate the sound with the image, even when the image is stilled, and whether a change in spatial location would affect this.

As in Experiment 5, I logged the time at which each utterance began and ended using used VideoCoder (1.2), a custom software tool designed for accurately time-stamping events in video. Gaze locations were then categorised according to which target was being fixated and whether that target was currently speaking. This log was prepared by the first author and

one other researcher to check reliability of the coding. Overall, averaged across all participants and six clip conditions, participants spent 38.87% of the time looking at a target who was currently speaking, 50.85% of the time on a non-speaking target and 10.28% of the time looking elsewhere. When comparing with results from Experiment 5's Control condition, these results are fairly consistent (see Figure 4.6). Table 4.5 shows this data split by condition. Figure 4.10 illustrates the average percentage looks on a current speaker. For further plots to illustrate these results see Appendix 4.

		Control		Freeze-Frame only		Freeze-Frame with change in spatial location	
% Fixations		Horizontal (1)	Stacked (2)	Horizontal (3)	Stacked (4)	Horizontal to Stacked (5)	Stacked to Horizontal (6)
On a speaking target	Average	44.40	42.69	38.29	36.35	37.37	37.99
	SD	16.55	18.43	17.80	12.93	12.87	16.52
On a non-speaking target	Average	48.94	52.93	48.85	56.65	50.23	45.66
	SD	15.80	16.36	15.50	14.31	13.54	15.18
Elsewhere	Average	6.66	4.38	12.86	7.00	12.40	16.35
	SD	16.22	11.82	13.32	8.78	11.01	14.63

Table 4.5. Average looks to targets currently speaking, a non-speaking target and elsewhere, split by condition.

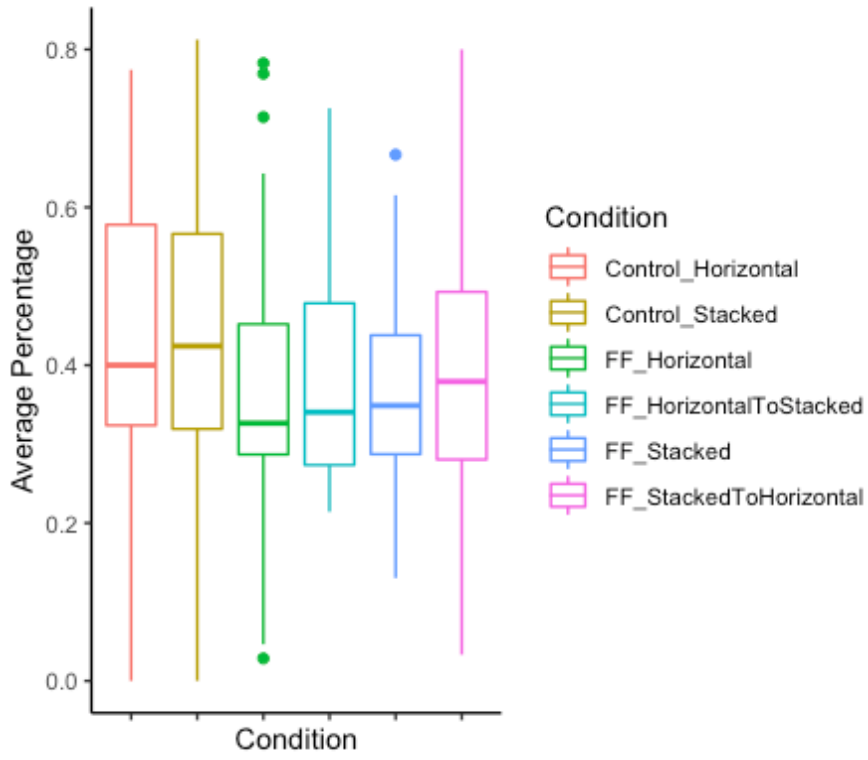

Figure 4.10. Demonstrates the average percentage of looks to speakers, split by condition.

When first observing at the average looks to a speaker, we can see there are decreases in percentages of fixations in all Freeze-Frame conditions, when comparing to the Control condition. However, a 1 x 6 repeated measures ANOVA found there was no significant effect of clip condition on looks to speakers, F (5,170) =1.52, p=.186. Equally, confirmatory pairwise comparisons revealed there were no significant differences between the conditions, hence participants looked to speakers similarly in all conditions. Therefore, when visually exploring the data, it seems there is a small but insignificant effect of freeze-framing the video when quantifying the amount of looks to targets currently speaking. Equally, when looking at our comparisons of interest (3 with 6 and 4 with 5), there were no significant differences. The means of these conditions are very similar demonstrating no effect of spatial change on looks to a current speaker.

I then explored whether the percentage of looks 'elsewhere' (not on a target) differed according to condition. A 1 x 6 repeated measures ANOVA established there was a

significant effect of condition, F (5,170) =5.57, p<.001. Bonferroni adjusted (SPSS) follow up tests established there were significant differences between the Stacked Control (2) condition with both spatial manipulation conditions (5) and (6). This demonstrated less looks to 'elsewhere' in the stacked control condition without any spatial change. The FF stacked condition (4) also significantly differed from the FF stacked to horizontal condition (6), with less looks elsewhere again without a spatial change. However, again, our targeted comparisons of interest showed no significant differences.

Discussion of Experiment 6

The present experiment used a unique paradigm to explore the ability of using audio only to follow conversation during a third-party group conversation. In line with previous research (Experiment 5), this experiment found that a large amount of visual attention was directed towards the targets. There were also similar shares of visual attention to targets dependent on target speakership. Overall, there were a number of similarities between the Control and manipulated conditions, which demonstrate a predisposition to look at target individuals in group settings, even when visual aspects are manipulated. This helps us to further understand gaze in complex group social settings.

This paper had two main aims. First, to explore whether Experiment 5's freeze-framed findings would be replicated. That being, that participants still show a preference for looking at targets and also whether there would be some evidence of conversation following. This experiment was specifically designed to test the robustness of this effect.

In the present experiment, we see similar results to those in the Freeze-Frame condition of Experiment 5. I demonstrate that participants look at targets 87% and 93% in the freeze-frame only conditions. Hence, the targets are clearly still capturing attention, even when there is no additional visual information to be gained by looking at a target. There are a

number of explanations for this finding. First, a low-level explanation taking a bottom-up approach is simply because the targets are the only area of interest on the screen and this may be present due to habit. In addition, the targets are a social element, which we reliably know captures attention (End & Gamer, 2019). Suggestions for future work to explore this further could be to include other non-related faces on screen to assess whether these areas compete for visual attention. If participants continue to look at the faces who were previously part of the fluid conversation, then it may suggest looking to the targets somehow facilitates conversational understanding. This is a potential second explanation, in that perhaps being able to pair the targets face with the audio of their voice helps to process the conversation.

In terms of whether the speaker was fixated at the time of their utterance, the results here also closely mimic results in Experiment 5, where a speaking target was fixated roughly 40% of the time in the Freeze-Frame condition. Although I cannot state this offers strong evidence on conversation following, there still seems to be some evidence of such. An idea to explore this further could involve using targets who are speaking a different language (which is not understood by the observer). Using this method, we could help tease apart whether looking to target speakers helps to comprehend the conversation, or whether we are drawn to the speaker nonetheless.

Secondly, I wanted to explore whether manipulating expected spatial location modulated this effect. In other words, would participants continue to look at the targets similarly even when the location of the freeze-framed targets moved? Overall, there were minimal differences in terms of targets being fixated and conversation following (looking at a speaker). There were slight reductions comparing the manipulated clips to the Control clips, which would be expected. However, the majority of these comparisons were non-significant and in particular, there were non-significant differences in the comparisons which were of interest. Therefore, in the present experiment, the small differences between the Control and

manipulated clips demonstrate how participants looked at speakers and non-speakers similarly throughout the clips. This is regardless of freeze-framing or the space in which they reside.

In terms of the theoretical implications of these findings, taken together the results suggest participants show a similar association for the visual and auditory components. In both experiments in this chapter, there isn't compelling evidence that we predominantly focus on a speaker (like we tend to do in a live situation (Ho et al., 2015). Instead, in these third-party situations we also seem to distribute our attention to other targets (supported by Foulsham and Sanderson, 2013). Explanations for this could be the complexity of the clips as stimuli. As previously identified, these clips are derived from an extremely naturalistic conversations, which may have impeded conversation following and the corresponding signalling cues which enable this. For example, it is a lot easier to follow a dyad scripted interaction (such as in Hirvenkari et al., 2013), where there is a more obvious moment of turn-taking. However, a higher-level explanation is that perhaps participants are more able to explore the scene and other target reactions to third-party conversations. By this I mean perhaps the social norms of looking at a person while they are speaking are not so internally enforced in a third-party situation and instead participants are more able to look at non-speaking targets as a form of social referencing. For example, it would be considered awkward to look at a non-speaker in a live social setting when another member of the group is speaking or 'taking the floor'. However, perhaps in the third-party situations described here, the participant feels freer to explore the reactions of the other individuals present. This would help explain any variability in comparing live and third-party observations (i.e., Laidlaw et al., 2013) and supports the modulating effects of social presence.

Overall, this research demonstrates the predisposition to look at targets even when audiovisual cues are manipulated. In both experiments, when movement is not present to

capture visual attention (i.e., without mouth movement and gestures) and in Experiment 5 when no sound was available, participants continue to fixate targets. In terms of theoretical significance, this indicates a strong association and demonstrates how we are able to utilize which cues *are* available to us (supporting research such as, Hirvenkari, et al., 2013 and Latif et al., 2018), in dynamic group settings. Whether this is due to conversation following or a visual preference for the targets needs some clarity. Study ideas to develop this further include using targets whom the participants are very familiar with, using targets speaking a different language and using targets rated with varying attractiveness. These ideas could help tease apart how visual attention is directed to conversation when audio and visual aspects are manipulated.

Chapter Summary

This chapter offered an exploration of visual and auditory modalities to facilitate gaze to speakers during conversation.

Experiment 5 investigated gaze during conversation in a realistic group of six individuals and in a more controlled laboratory study where third-party observers watched videos of the same group. In both contexts I explored how gaze allocation is related to turn-taking in speech. Gaze behaviour in the real, interactive situation was similar to the fixations made by observers watching video. This provides support for using third-party video observations to explore social attention in experiments within this thesis and beyond.

In the third-party observers, experimental video clips were edited to either remove the sound, freeze the video or transition to a blank screen, allowing us to determine how shifts in attention between speakers depend on visual or auditory cues. Eye tracked participants often fixated the person speaking and shifted gaze in response to changes in speaker, even when sound was removed, or the video freeze-framed. These findings suggest we sometimes fixate the location of speakers even when no additional visual information can be gained. Hence, in the third-party observers I manipulated the availability of both visual and auditory elements of the clips, and therefore the accompanying signalling behaviours, with participants using the audio-visual counterparts to attend to the conversation.

Experiment 6 further explored the strength of this sound-image association in the freeze-frame condition when the target's location is spatially manipulated. The results replicate findings of Experiment 5 and additionally explored under which conditions this gaze pattern occurs. I questioned to what extent the expected spatial location affects fixations to target speakers with the addition of freezing and manipulating the target's position on screen. To my knowledge, this is a novel approach to studying the strength of association from the sound and source of speakers in a group setting. Overall, looks to frozen and

spatially manipulated videos were very similar compared to control clips in terms of looking to targets and looking to targets currently speaking. This implies there is a strong association between target voices and frozen images. This research replicates previous work in Experiment 5, but also adds an additional novel element of spatial location manipulations.

Altogether, the results presented in Chapter 4 use a novel approach to offer both a comparison of interactive and third-party viewing and the opportunity for controlled experimental manipulations. This delivers a rich understanding of gaze behaviour and multimodal attention during conversation following. The following chapter explores the manipulation of visual signals, this time in live situations.

Chapter 5: The effects of clinical traits of ASD and ADHD on viewing behaviour during conversation watching

Parts of this chapter (Experiment 7 and 8) were produced with colleague Miss A.P Cedillo-Martinez.

This chapter investigates how gaze behaviour varies in populations with atypical traits, with the addition of manipulating the presence of the eyes. Three experiments are presented which explore the effect of these traits on eye movements. Using the same video stimuli in all experiments (Experiments 7, 8 and 9), I investigate gaze behaviour and conversation understanding, and how the performance of both corresponds with high and low traits in these clinical disorders.

Experiment 7 & 8 - Occluding eyes in conversation: the effects on gaze following in ADHD and ASD- like traits

Experiment 7 and 8 were co-produced with colleague Miss A.P Cedillo-Martinez. JD was responsible for idea conceptualization, building the experiments and collecting the data. Analysis and manuscript preparation was shared equally between JD and AC but restructured by JD for this thesis.

Social cues facilitate our social interactions and communication. As described in Chapter 1, atypical social interactions are present in the behaviours of individuals with Autism Spectrum Disorder (ASD) as well as Attention Deficit Hyperactivity Disorder (ADHD). One suggestion is that these clinical populations do not utilise social cues in a way that neurotypical populations do. However, surprisingly little empirical work has been devoted to investigating the effect of concealing such social cues during a conversation. The aim of Experiment 7 and 8 is to examine visual attention to a pre-recorded natural group conversation while individuals within those scenes had their eyes occluded (using sunglasses). Additionally, to further understand how populations with traits of ASD and ADHD utilise the eyes as a social cue, we used participants of low and high traits of ASD (Experiment 7), and ADHD (Experiment 8).

Introduction to Experiment 7 and 8

In Experiment 7 and 8, we collect eye tracking data from a subclinical sample of ASD and ADHD. A subclinical sample refers to those displaying symptoms of a specific disorder but who do not meet the criteria for the disorder per se. Interestingly, this subclinical sample might facilitate the understanding of the disorders, since it is less likely that this sample have taken any psychostimulant during their life course and/or have been under any clinical intervention.

Here, we investigate the effect of occluding the eyes when observing a pre-recorded natural conversation. We further aim to understand how populations with high and low traits of ASD (Experiment 7) and ADHD (Experiment 8) utilise the eyes as a social cue. We use a similar methodology to Dawson and Foulsham's (2021) study, whereby participants will watch video clips depicting target individuals sitting around a table engaging in a group discussion. This methodology comprises of third-party participants watching group conversations which have previously been recorded, with clips prepared for a static eye-tracker. In half of the clips individuals in the scene (targets) will be wearing sunglasses to occlude their eyes (Sunglasses condition) and in the remainder, their eyes will be visible (Control condition). We have three main objectives.

First, we explore how occluding the eyes affects fixations to individuals within the scene. We investigate to what extent overall looking to people and their facial features (eyes and mouths) are affected by sunglasses occluding their eyes. Previous research has reliably found that we tend to look at social aspects of the scene (Flechsenhar, Rösler & Garmer, 2018) and in particular the eyes (Friesen & Kingstone, 1998). However, the role of the eyes in conversation is not always clear, as it is rare the eyes are observed alone without accompanying head movements and gestures. One way to examine the role of the eyes in

social attention, is by occluding them (such as with sunglasses). To our knowledge, there is limited research exploring the effect of occluding the eyes during conversation. Perhaps the most relevant work on the effects of eyes occlusions can be exhibited in studies of face recognition (Hockley, Hemswoth & Consoli, 1999), facial expression processing (Roberson, et al., 2012), emotion identification (Zhang, Tjondronegoro & Chandran, 2014) and cooperative tasks (Hanna & Tanenhaus, 2004; Metzing & Brennan, 2003; Boucher et al., 2012). In these studies, occluding the eyes with sunglasses impeded typical social processes, which leads us to question how this would affect attention to dynamic conversation.

For this reason, we expect Control (low trait) participants to reliably look to the target individuals within the scene and to the eyes when they are visible. When occluding the eyes with sunglasses, we may expect a decrease in looks to the eyes, as there is no additional benefit (i.e., no understanding of intentions or signalling) to be gained by fixating this area. Equally, we may see no difference due to habit or even an increase in attention as this is a novel item within the scene.

Second, we assess how and when a speaker is observed when occluding the eyes with a comparison between the Control and Sunglasses conditions. We therefore test the effect of using the eyes as a signalling cue. In a typical population, the majority of fixations to third-party video tend to be directed toward the current speaker. The ability to monitor a conversation and direct gaze to an appropriate person might depend on gaze following of the individuals within the scene. Hence, we will explore whether wearing sunglasses impedes the observer's ability to follow turn-taking conversation. This will be investigated in terms of frequency of looking behaviour to those currently speaking and in a time-based analysis to assess the moment at which a speaker is fixated (conversation following).

Finally, we assess the prior objectives and additionally compare findings for high and low traits of ADHD and ASD. ASD literature has been more widely studied in terms of social interaction and communication impairments than ADHD. However, in both disorders there appears to be visual discrepancies, which can serve as an essential factor for deeper understanding. We are therefore interested in how high and low trait individuals differ in viewing people and whether occluding the eyes (using sunglasses) affects visual attention. We have previously discussed that looking to the eyes of others is an automatic response. If those with high traits of ASD and ADHD have an avoidance response, then occluding the eyes may facilitate looks to this area, in comparison to when the target individual's eyes are visible. In line with previous literature, we expect fewer fixations to the eye area for the high trait groups in the Control condition, particularly for the high trait ASD population. As there is limited evidence for social attention differences in ADHD, this is more exploratory in nature. Additionally, as it suggested these populations do not follow typical signalling cues, we expect there may be disparities in terms of when targets are fixated (conversation following) when eyes are visible or occluded.

Materials and methods

The aims and analysis of this experiment were pre-registered (see https://osf.io/c3jvk/).

Apparatus

The same type of apparatus and stimuli were used for Experiment 7 and 8. Eye position was recorded using the Eyelink 1000 (SR Research), a video-based eye-tracker that samples pupil position at 1000Hz. A nine-point calibration and validation procedure ensured all recordings had a mean spatial error of better than 0.5 degrees. Head movements were restricted using a chin rest and sound was played through headphones. Participants sat 50 cm away from the monitor so that the stimuli subtended approximately 30°x17° of visual angle at

1024*576 pixels. Saccades and fixations were defined according to Eyelink's acceleration and velocity thresholds.

Stimuli

The video clips shown to participants depicted six individuals (referred to as targets) having a discussion while sitting around a table, with only 3 individuals (one side of the table (T1-T3)) in view in each clip.

The clips were derived from a 1 hour recording of four group conversations as in Experiment 5 and 6 (see Experiment 5 Methods for details). In addition, the targets were also given sunglasses to wear for an equal proportion of the recording. The targets were given an equal number of questions to discuss (with and without sunglasses), and the conditions were counterbalanced in terms of time and questions. When given the sunglasses, targets were under the impression that this was to examine how they behave when wearing them during conversation. The experimental clips were selected from each continuous recording and featured moments where all visible targets spoke at least once. These clips were selected to ensure that Targets 1, 2 and 3 were the predominant speakers, with minimal involvement from the targets on the other side of the table.

Two 35 second clips were chosen for each group, one with and one without sunglasses. The result was 8 experimental clips. Half of these clips included a conversation with the eyes present (Control condition) and half of the clips featured a conversation with the eyes occluded (Sunglasses condition), an example of the scene from the clips is shown in Figure 6.1.

Figure 6.1 An example video frame from the Control and Sunglasses condition, showing the three targets.

Participant procedure

The eye tracked participants read and completed consent forms and were asked to confirm that they had normal or corrected to normal vision before beginning the experiment. After the participant's right eye had been successfully calibrated and validated, the experiment began. There were 8 experimental trials (4 Sunglasses and 4 Control clips). Trials were presented in a randomised order. After each clip, questions were presented which asked about the events in the video clip. (e.g., "which person was wearing a green t-shirt?"), and participants responded by pressing '1', '2', or '3' on the keyboard to indicate one of the three targets. The questions were piloted for difficulty before beginning the experiment and only used to ensure participants were paying attention to the clips. Participants were given verbal and written instructions regarding the experimental procedures. The experiment took a total of approximately 10 minutes.

Experiment 7 (ASD)

Experiment 7 was designed to examine eye movements in participants with ASD-HT (high trait) and ASD-LT (low trait) whilst watching conversation videos, with the additional manipulation of occluding the target's eyes.

Participants

Students from the University of British Columbia were recruited to take part, based on their ASD-10 questionnaire responses from a larger sample. We collected eye-movement data from 41 individuals. The ANOVA interaction with 40 participants across two conditions would be sensitive to effects of $\eta^2 = .04$ with 80% power (alpha = .05). All of the participants reported normal or corrected-to-normal vision. Participants were granted with course credit for their participation. The study was approved by the ethics board at the University of British Columbia.

ASD classification

In pre-screening, all psychology students at UBC were asked to complete the AQ10 questionnaire (Allison et al., 2012) when beginning the semester. Participants were then selected from this population. We used the symptom checklist to classify participants with high traits of ASD (ASD-HT) and low traits (ASD-LT). To do so, we considered the total score obtained in the entire questionnaire. Participants were classified as ASD-LT where they had a total score of less than 2 and ASD-HT where they had a score of 6 and above. We invited participants to the lab only if they met these criteria, with those who responded scheduled for testing until our sample size was met. In the ASD–HT group (N=21) there were 15 females, (mean (SD) age of 19.86 (1.85) years), and in the ASD–LT group (N=20) there were 19 females (mean (SD) age of 21.25 (1.02) years).

Analysis and results

No participants were excluded for poor calibration. For this reason, 41 participants were included in the main analysis.

General viewing behaviour

First, we examined how participants with ASD-HT and ASD-LT responded to the conversation clips by analysing general eye movements as presented in Table 6.1. This was

investigated to determine whether there were any global differences in the way the two groups moved their eyes.

	Mean Fixations Per Clip	Mean Fixation Duration (ms)
ASD-HT	76.82	388.43
SD	*12.51*	*83.19*
ASD-LT	84.43	361.10
SD	*16.31*	*97.35*

Table 6.1. Number of fixations per clip, and fixation duration (in milliseconds) averaged for each group.

To establish if there were any differences between the participant groups, we ran two independent sample t-tests for average number of fixations, (t (39) = 1.69, p =.10 d= 0.26), and average fixation duration (t (39) =-0.96, p =.33, d = 0.15). These both revealed non-significant differences between the two groups, indicating that participants' general viewing behaviour was similar.

How are targets fixated?

We then analysed how each group fixated the targets in the scene, in each condition, both overall and when that target was speaking.

Fixations to targets

We defined a region of interest (ROI) around each target individual as in Experiment 5. The static regions of interest (ROIs) were drawn around each of the 3 whole targets using SR Research Data Viewer, with results shown in Table 6.2.

		Mean % Fixations to Targets			
		Control		Sunglasses	
		Targets	Elsewhere	Targets	Elsewhere
ASD-HT	M	97.83	2.17	98.04	1.96
	SD	*1.71*	*1.71*	*3.07*	*3.07*
ASD-LT	M	97.73	2.27	98.45	1.55
	SD	*1.42*	*1.42*	*1.95*	*1.95*

Table 6.2. Represents the average percentage of fixations to targets split by Condition and Group.

Pooling these ROIs together, it was clear that participants directed almost all of their fixations to the people in the scene. The average percentage of fixations on the ROIs was entered into an analysis of variance (ANOVA) with the within-subjects factor of Condition (Sunglasses or Control) and the between-subjects factor of Group (ASD-HT or ASD-LT). There was no effect of condition ($F<1$), or group ($F<1$), and no interaction between condition and group ($F<1$). This suggests that both groups and conditions behave similarly when analysing overall looks to targets.

Fixations to targets' eyes and mouth

Previous studies have found a tendency to fixate the eyes in a general population in both images and video (e.g., Birmingham, Bischof & Kingstone, 2007; Klin et al., 2002). For this reason, we investigated whether there was an effect of Condition and Group on looks to specific regions of the face. Moving ROIs were drawn around the eyes and mouth of each of the 3 targets using Data Viewer (SR Research). The positions throughout the recordings were adjusted by slowly playing the clip back with 'mouse record' (an inbuilt function in Data Viewer), which allowed the tracking of these areas when targets moved. Fixations on targets were then analysed to determine whether they were inside the total target area. Table 6.3 shows the average percentages of fixations on the eyes, mouth and elsewhere on the target.

		Mean % Fixations to Targets					
		Control Condition			Sunglasses Condition		
		Eyes	Mouth	Elsewhere	Eyes	Mouth	Elsewhere
ASD-HT	M	25.27	14.29	60.44	41.47	15.94	42.59
	SD	*15.77*	*13.42*	*13.05*	*21.77*	*11.72*	*19.48*
ASD-LT	M	32.24	17.38	50.37	45.59	23.17	31.24
	SD	*15.25*	*14.08*	*8.08*	*20.00*	*15.46*	*15.58*

Table 6.3. The mean percentage of fixations to targets' eyes and mouth and elsewhere, split by Group (low and high traits of ASD) and Condition (Control and Sunglasses). Fixations outside the main target ROIs are not included here.

Participants' average probability of fixations to the ROIs were entered into an ANOVA with the within-subject factors of condition (Sunglasses and Control), area (mouth and eyes) and the between-subjects factor of group (ASD-HT or ASD-LT). There was an effect of area (F (1,39) =17.52, p<.001, η^2 =.10), indicating that participants looked more to the eye area compared to the mouth area. There was an effect of group (F (1,39) =6.27, p=.017 η^2 =.14), indicating that the ASD -LT group made more fixations to both areas (eyes and mouth) in comparison to the ASD -HT group. There was also an effect of condition, (F (1,39) =122.39, p<.001, η^2 =.76). Interestingly, this was qualified by an interaction between condition and area (F (1,39) =29.80, p<.001, η^2 =.43), indicating that the bias to look at the eyes rather than the mouth was more pronounced in the Sunglasses condition compared to the Control condition. There were no other significant interactions (F<1).

Are speakers fixated more?

Fixations to speakers

We then analysed the looks to targets who are currently speaking. As in Experiment 5, to accurately log when the utterances began and ended, we used the auditory signal with the visual signal to assist in identifying the speaking target. For this we used VideoCoder (1.2), a custom software tool designed for accurately time-stamping events in video. Gaze locations were then categorised according to which target was being fixated and whether that

target was currently speaking. The average percentage of fixations to speaking targets can be seen in Table 6.4.

		Mean % Fixations			
		Control Condition		Sunglasses Condition	
		Speakers	Elsewhere	Speakers	Elsewhere
ASD-HT	M	45.33	54.67	53.19	46.81
	SD	*6.49*	*6.49*	*4.27*	*4.27*
ASD-LT	M	48.21	51.79	54.91	45.09
	SD	*4.86*	*4.86*	*3.98*	*3.98*

Table 6.4. The average percentage of fixations on speaking targets split by Condition and Group. The elsewhere category includes fixations on the other non-speaking targets and any non-target fixations.

Participant means were entered into an ANOVA with the within-subjects factors of Condition (Sunglasses and Control), and between-subjects factor of group (ASD-HT or ASD-LT). There was an effect of condition (F (1,39) = 139.01, p<.001, η^2 = .78), indicating that participants made more fixations to speaking targets in the Sunglasses condition than in the Control condition. There was no effect of group (F<1), and no interaction between condition and group (F<1).

When are speakers fixated?

We then analysed at which point in time participants made a fixation to a speaker, to understand whether conversation following varied for participant group or clip condition. The start times of each utterance (taken from each target in each clip) were used to create 10ms bins ranging from -1000msec pre speech beginning and 1000msec post utterance beginning. We then compared these bins to the fixation data and coded bins as to whether they contained a fixation on a target speaking, a fixation elsewhere or no fixation. We extracted a percentage of looks to a speaker for each bin (averaged across participant, Condition and the multiple utterances), which we could then compare within the time period of interest. The result was an estimate of the probability of looking at a speaker, time-locked to the beginning of their

speech. We have previously shown that participants sometimes move in advance of the change in speaker, and that the time course is affected by both auditory and visual information (Dawson & Foulsham, 2021).

Probability of Fixations

We then analysed the probability of fixations to a target speaker upon the utterance beginning. This analysis was calculated across all data points from the 201 time frames (10msec bins -1000 to +1000msec) from each participant and each condition giving a total of 16,482 data points to analyse. A visual example of the probability of fixations landing on the target speaker, relative to when they started speaking across the analysed time course can be seen in Figure 6.2. This graph shows how the relative frequency of looks to a target increases around the time that they begin talking – a pattern seen in both groups and both conditions.

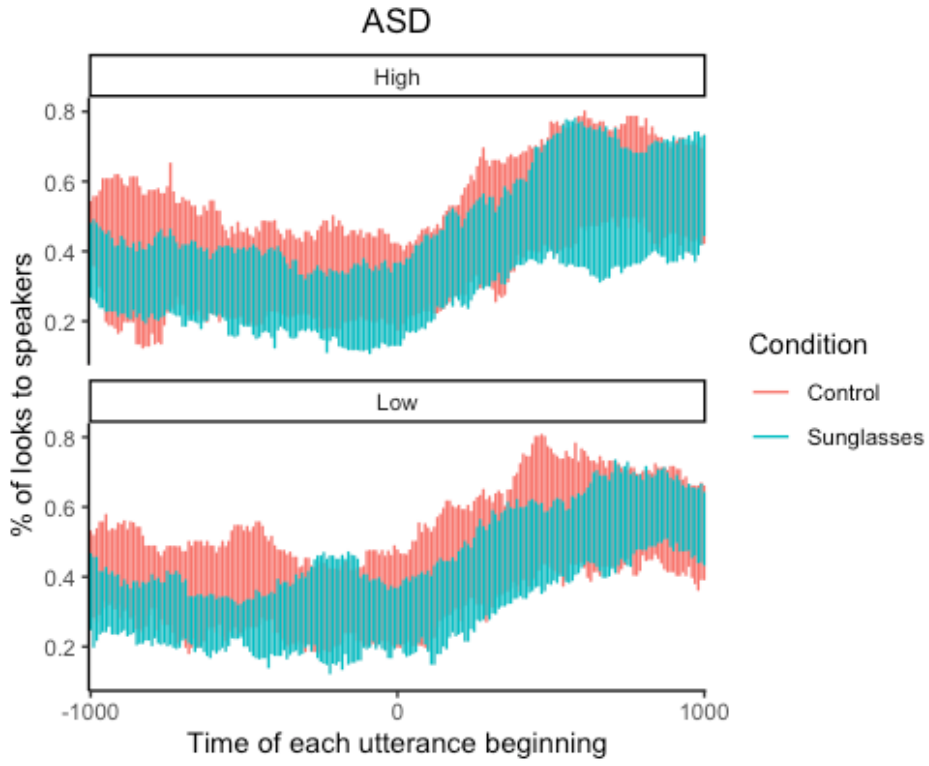

Figure 6.2. Probability of fixations being on the speaker, relative to when they started speaking averages across condition and group. A time of 0 indicates the time at which a speaker began speaking.

To analyse the overall probability of fixations within this time frame we calculated summary averages for Condition and Group, demonstrated in Table 6.5. A mixed ANOVA established there was a significant effect of Condition (F (1,39) =92.574, $p < 0.001$), indicating more looks in the Control condition. There was no significant effect of ASD group (F (1,39) =0.101, p=0.753) and a non-significant interaction F (1,39) =0.614, p= 0.438).

		Mean Probability of Fixations	
		Control	Sunglasses
ASD-HT	M	43.7	38.1
	SD	*13.7*	*14.3*
ASD-LT	M	43.9	37.4
	SD	*12.2*	*12.6*

Table 6.5. Shows the average probability of a fixation (from -1000msec to +1000msec of utterance).

Highest Percentage Bin

Second, we were interested in at which point in time (-1000msec to +1000msec post utterance beginning) the participant was most likely to be looking at a speaker. We analysed this by finding the time bin with the maximum percentage for each participant. An average of those times can be seen in Table 6.6. This is also demonstrated visually in Figure 6.3.

		Maximum Bin	
		Control	Sunglasses
ASD-HT	M	735.24	806.67
	SD	*154.75*	*137.56*
ASD-LT	M	724.50	803.50
	SD	*183.58*	*156.18*

Table 6.6. The time at which participants were most likely to be looking at a speaker, averaged across participants and conditions.

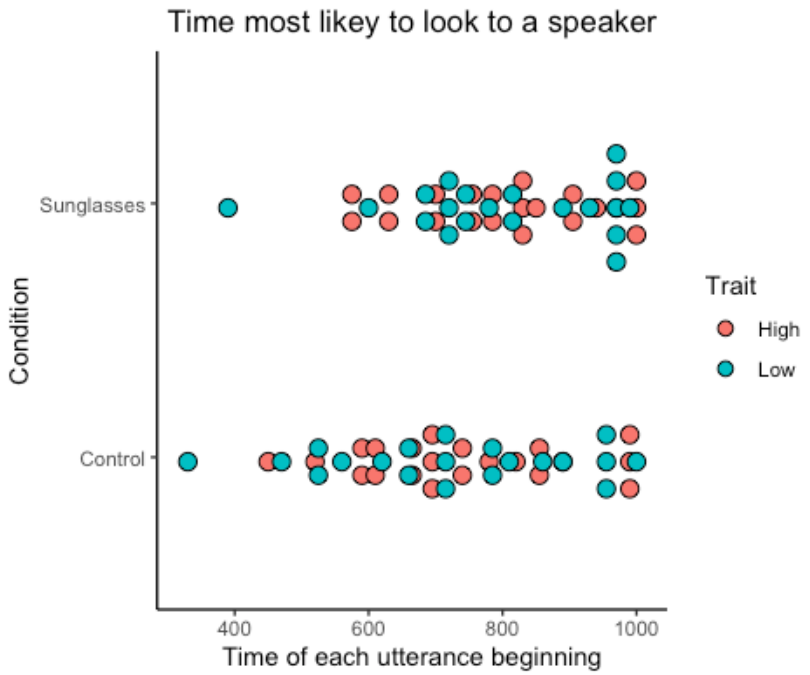

Figure 6.3. The average time at which participants were most likely to be looking at a speaker, split by condition and ASD group. A time of 0 would indicate the time of the speech beginning.

A mixed ANOVA established there was a non-significant effect of Group ($F<1$). There was a significant effect of Condition ($F\ (1,39) = 5.22$, $p<.05$) and a non-significant interaction ($F<1$). Therefore, in the Control condition, compared to the Sunglasses condition, participants were more likely to look at a target earlier when their utterance began. The ASD group had no effect.

Overall, looks were more likely and earlier in the Control condition compared to Sunglasses condition when looking at this time frame. This result is different to the percentage of overall looks to speakers, where there was more time spent on the speakers in the Sunglasses Condition. Hence, there appears to be a shift in attention over time.

Experiment 8 (ADHD)

As in Experiment 7, Experiment 8 was designed to examine participants' eye movements whilst watching conversation videos, with the manipulation of occluding the target's eyes. However, this study uses participants who fall into ADHD-HT (high trait) and ADHD-LT (low trait) groups.

Participants

After pre-screening 248 students, we collected eye movement data from 40 individuals who were students from University of Essex. All of the participants reported normal or corrected-to-normal vision. Participants were granted with five pounds for their participation. The study was approved by the ethics board of the University of Essex.

ADHD classification

In the pre-screening, 248 participants were asked to complete the Adult ADHD Self-Report Scale (ASRS; Kessler at al., 2005) via online using Qualtrics. We used the symptom checklist to classify the participants with high traits (ADHD-HT) and low traits (ADHD-LT) of ADHD. Scores on the ASRS checklist varied from 13 to 61 and the mean (SD) score was 37.87 (12.50). We considered the total score obtained in the entire questionnaire and chose the highest (between 45 to 61) and lowest (between 13 to 34) scores for high and low traits, respectively. From those participants who completed the ASRS, we found 25 of those reported with high traits and 49 reported as low trait. We contacted 20 participants from each group and if we did not have a response, we randomly selected another participant.

We invited participants to the lab only if they met these criteria. In the ADHD-HT group (mean (SD) age of 21.15 (1.48) years) there were 10 females. Two participants reported being diagnosed with ADHD, one with dyslexia, and one with depression. In the ADHD-LT group (mean (SD) age of 22.6 (3.87) years) there were 17 females. 1 participant

reported be diagnosed with Generalised Anxiety Disorder. No participants reported taking psychostimulants or reported as diagnosed with ASD or ADHD.

Analysis and results

The analysis in Experiment 7 was repeated for this data. We excluded one participant due to poor calibration, this participant was classified as ADHD-HT. For this reason, 39 participants were included in the main analysis.

General viewing behaviour

First, we examined how participants in the ADHD-HT and ADHD-LT groups responded to the conversation clips by analysing general eye movements as presented in Table 6.7. We included this analysis to understand whether clips were overall visually attended to differently between groups.

	Mean Fixations Per Clip	Mean Fixation Duration (ms)
ADHD-HT	78.15	390.40
SD	*13.27*	*65.11*
ADHD-LT	78.52	397.83
SD	*11.16*	*57.22*

Table 6.7 Mean fixations per clip and fixation duration (in milliseconds) averaged for each group (ADHD-HT and ADHD-LT).

We ran two independent sample t-tests on mean number of fixations, (t (37) =-0.09, p =.92, d=0.02), and mean fixation duration (t (37) =-0.37, p =.70, d= 0.06). These both revealed non-significant differences between the two groups, indicating that participants' general viewing behaviour was similar (as in Experiment 7).

How are targets fixated?

Fixations to targets

As in Experiment 7, to explore how the participant group and clip condition affected how much participants looked at the targets, we defined a region of interest (ROI) around each target individual. See Experiment 7 for further details.

Pooling these ROIs together, again, participants fixated the people in the scene most. The average percentage of fixations on the ROIs was entered into an ANOVA with the within-subjects factor of Condition (Sunglasses or Control), and the between-subjects factor of Group (ADHD-HT or ADHD-LT). There was no effect of Condition (F<1), or Group (F<1) and no interaction (F<1). These results suggest that both groups behave similarly when looking at the targets with and without sunglasses (see Table 6.8).

		Mean % Fixations			
		Control	Elsewhere	Sunglasses	Elsewhere
ADHD-HT	M	97.28	2.72	98.69	0.68
	SD	*2.68*	*2.68*	*1.43*	*0.75*
ADHD-LT	M	98.01	1.99	98.69	0.68
	SD	*2.07*	*2.07*	*1.43*	*0.75*

Table 6.8. Represents the average percentage of fixations to targets, split by Condition and Group.

Fixations to targets' eyes and mouth

As in Experiment 7, moving regions of interest (ROIs) were drawn around the eyes and mouth of each of the 3 targets. Fixations on targets were then analysed to determine where they were inside the area. Table 6.9 shows the average percentage of fixations to targets' eyes and mouth and elsewhere on the target throughout the clips.

		Mean % Fixations to Targets					
		Control			Sunglasses		
		Eyes	Mouth	Elsewhere	Eyes	Mouth	Elsewhere
ADHD-HT	M	25.45	16.03	58.52	26.21	16.49	57.30
	SD	*18.28*	*15.48*	*16.15*	*18.16*	*14.77*	*16.69*
ADHD-LT	M	24.62	21.01	54.38	26.05	19.89	54.06
	SD	*17.39*	*14.39*	*18.32*	*16.77*	*14.12*	*18.53*

Table 6.9. The mean percentage of fixations to targets' eyes, mouth and elsewhere on the target split by Group and Condition. Fixations outside the main target ROIs are not included here.

Participants' average probability of fixations were entered into an ANOVA with within-subject factors of condition (Sunglasses or Control), ROI (eyes and mouth) and the between-subjects factor of Group (ADHD-HT or ADHD-LT). There were no effects or interactions (all $F<1$), indicating eyes and mouth regions were fixated similarly regardless of Group and Condition. This pattern was slightly different from Experiment 7, where the addition of sunglasses led to more looks at the facial regions.

Are speakers fixated more?

Fixations to speakers

We then analysed the looks to targets who are currently speaking. The same speaking log was used as in Experiment 7 to analyse when targets were speaking and whether the participant observer was looking at the speaker. The average percentage of fixations to speaking targets can be seen in Table 6.10.

		Mean % Fixations to Speaking Targets			
		Control		Sunglasses	
		Speakers	Elsewhere	Speakers	Elsewhere
ADHD-HT	M	45.59	54.41	51.65	45.11
	SD	5.85	5.85	6.56	6.35
ADHD-LT	M	46.77	53.23	53.24	49.57
	SD	4.70	4.70	5.97	5.11

Table 6.10. The average percentage of fixations on speaking targets split by Condition and Group. The elsewhere category includes fixations on the other non-speaking targets and any non-target fixations.

Participant averages were entered into an ANOVA with the within-subjects factors of Condition (Sunglasses and Control), and the between-subjects factor of Group (ADHD-HT or ADHD-LT). There was an effect of Condition (F (1,37) = 41.38 p= 0.001, η^2 = 0.53). There was no effect of group (F<1) or condition and group interaction (F<1), demonstrating participants looked more to speakers in the sunglasses condition regardless of their group. This replicates the pattern observed in Experiment 7.

When are speakers fixated?

As in Experiment 7, we then analysed at which point in time participants made a fixation to a speaker. Using the same method, we compared the time log of each targets' utterance with whether there was a fixation on that speaking target (plus and minus 1000msec). The result was an estimate of the probability of looking at a speaker, time-locked to the beginning of their speech.

Probability of fixations

As in Experiment 7, we analysed the overall average probability of fixations to a speaker within a -1000 to +1000msec time frame for each target's utterance. This analysis across all data points from the 201-time frames, from each participant and each condition gave a total of 15,678 data points to analyse. A visual example of the probability of fixations being on the speaker, relative to when they started speaking across the analysed time course can be seen in Figure 6.4. The pattern is similar in both experiments.

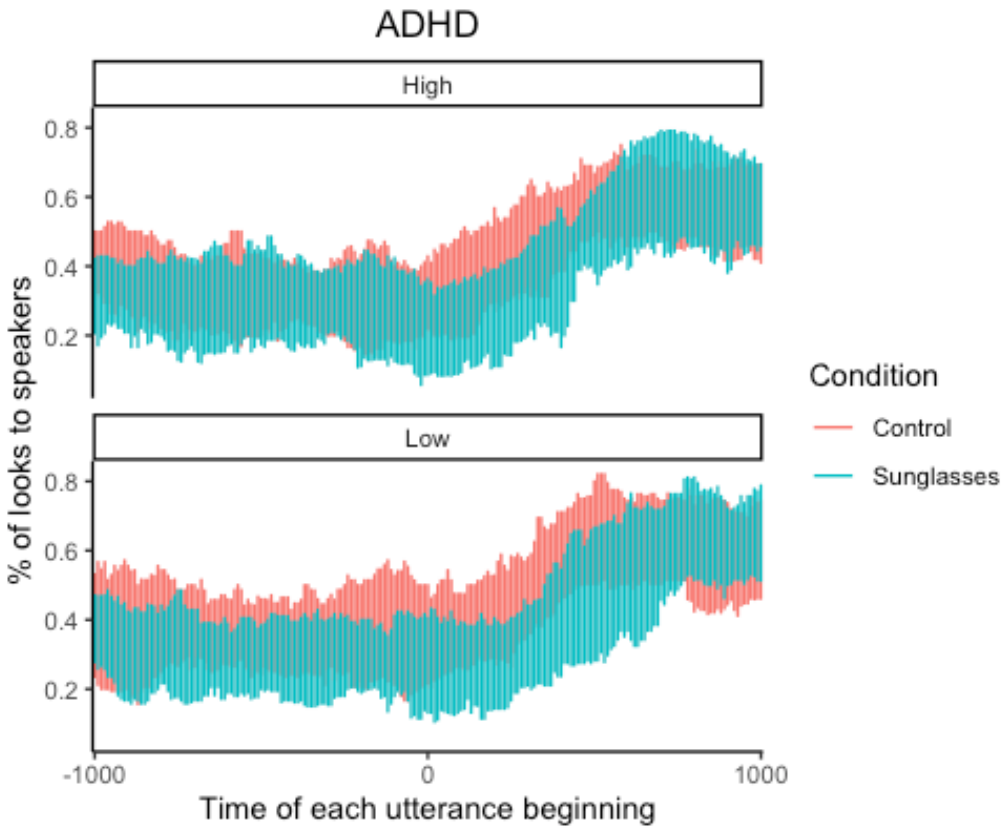

Figure 6.4. Probability of fixations being on the speaker, relative to when they started speaking averages across Condition and Group. A time of 0 indicates the time at which a speaker began speaking.

To analyse the overall probability of fixations within this time frame we calculated summary averages for Condition and group, demonstrated in Table 6.11. A mixed ANOVA established there were significant differences between ADHD groups (F (1,37) =5.231, p= 0.028), Conditions (F (1,37) =107.993, p<.001) and a significant interaction (F (1,37) =4.966, p=0.032). p<.001). Pairwise comparisons with a Bonferroni correction demonstrated there were significant differences between the Condition for both ADHD groups (p<.05). This demonstrates there were significantly more looks within this timeframe in the Control condition, with more looks for the ADHD-LT group. This is interesting, given in the same analysis with ASD-LT and ASD-HT, there were no significant differences.

		Control	Sunglasses
ADHD-HT	M	42.3	37.2
	SD	*13.4*	*15.0*
ADHD-LT	M	45.4	37.5
	SD	*13.3*	*14.3*

Table 6.11. Shows the average probability of a fixation (from -1000msec to +1000msec of utterance).

Highest percentage bin

Second, we were interested in quantifying at which point in time (from -1000msec to +1000msec post utterance beginning) the participant was most likely to be looking at a speaker. We analysed this by finding the maximum percentage for each 10msec bin for each participant. An average of those times can be seen in Table 6.12. This is also demonstrated visually in Figure 6.5.

		Maximum Bin	
		Control	Sunglasses
ADHD-HT	M	690.00	776.32
	SD	*156.95*	*123.12*
ADHD-LT	M	661.50	822.00
	SD	*147.73*	*124.50*

Table 6.12. Shows the time at which participants were most likely to be looking at a speaker, averaged across participants and conditions.

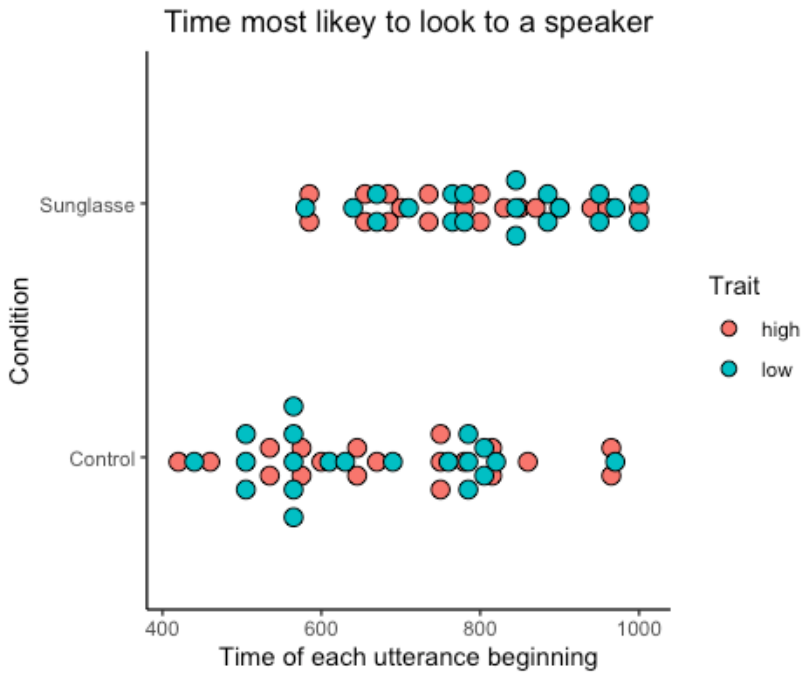

Figure 6.5. The average time at which participants were most likely to be looking at a speaker, split by Condition and Group.

A mixed ANOVA established that there was a non-significant effect of Group (F<1). There was a significant effect of Condition, (F (1,37) =13.82, p<.001), and a non-significant interaction (F<1). Therefore, in the Control condition, when target eyes were visible, participants were more likely to look at a target earlier when their utterance began. The participants' ADHD group had no effect.

Overall, as in Experiment 7, looks were more likely and earlier in the Control condition compared to Sunglasses condition when looking at this time frame around targets beginning their utterance.

Between experiment comparison

Overall, the analysis of Experiment 7 and 8 explored how participants looked to targets, how they looked to targets who were currently speaking and the timing at which this conversation following occurred. In Experiment 7 (ASD) those with high traits looked less to targets eyes and mouth, in line with expectations from previous literature. In Experiment 8

(ADHD), there were minimal differences in our analysis questions in terms of participants with high and low traits. Interestingly, in both experiments, conversation following seemed to be facilitated by visible eyes. Participants were slower and less likely to fixate a speaking target upon speech beginning than when the targets' eyes were covered by sunglasses. This was also more prevalent in ADHD-LT than ADHD-HT. These behavioural differences when occluding the eyes, was irrelevant of group in Experiment 7 (ASD). Next, we discuss the key findings and the implications.

Discussion of Experiment 7 and 8

Experiments 7 and 8 provided an innovative way to explore the effect of occluding the eyes on conversation following and additionally assess any differences this has on high trait and low trait ADHD and ASD populations. Interestingly, we found no differences in looking to the social areas (targets as a whole), with around 99% of fixations on targets in all trials. However, when we broke this down further into looking at speaking target, looks to specific regions and the timing of gaze following, there were a number of interesting findings. Collectively these findings highlight the complexities of how gaze is distributed in group settings with and without the eyes as a signalling cue for different populations with neurodivergent traits.

How are targets fixated?

In both experiments, we first compared the proportion of looks to targets as a whole. In both clip conditions and both participant groups, we saw extremely high percentages to targets in general. This is in line with previous literature of attention to social aspects (e.g., Flechsenhar, Rösler & Garmer, 2018; End & Gamer, 2019; Dawson & Foulsham, 2021) and is not surprising considering the targets were the only moving and social element within the scene. However, we may have expected the ADHD and more likely, the ASD group to show

a slight decrease in fixations to the social stimulus (as in Freeth et al., 2013 and Ristic et al., 2005). Arguably the results in the present study could be due to fact the targets collectively take up a large proportion of the screen. However, the proportions are not near these limits. Perhaps with a non-social competing object within the scene, results may have reflected a more diverse pattern. Another suggestion could be that high trait participants show less of an avoidance response in third-party viewing, where they are not actively engaging in conversation. For example, perhaps this population are more able to explore the scene without any implied or explicit social presence.

We were then particularly interested in looking behaviour to targets' eyes and mouth. Previous research (such as Ristic et al., 2005 and Serrano et al., 2018) indicates there could have been differences in visual attention to the eyes for high trait populations. When we split the fixations to targets into specific regions (eyes and mouth) in Experiment 7 (ASD) we saw group differences, with the low trait population looking more to the eyes and mouth than the high trait group. This supports previous research such as by Freeth et al., (2013) and Klin et al., (2002), but in addition demonstrates this effect is present in a larger group setting rather than static or more simplistic conversation stimuli. These differences were not present in the ADHD group.

Interestingly, in all populations, on average, there were more looks to the eyes in the Sunglasses condition than the Control condition. Previous research reports the prevalence of looking to the eyes (e.g., Birmingham, Bischof & Kingstone, 2007). However, in this experiment participants continued to look to the eye area in the Sunglasses condition, despite not being able to view the eyes to gain information. Explanations for this include that it is a habit or a novel aspect of the scene.

Are speakers fixated more?

We then explored how looking behaviour varied to targets who were currently speaking. In both groups and conditions, we saw around 50% of fixations to a target who is currently speaking. Previous research in dyad pairs indicate higher percentages of looks to speakers (i.e., Argyle & Ingham, 1972). However, the present research reflects attention decisions in a group interaction (with results similar to that of Dawson and Foulsham, (2021)). The group dynamic means the lower reported percentages are not surprising given the additional attention decisions required by the participants.

Again, in both experiments we found there were more looks to speakers when targets were wearing sunglasses. As discussed, this may be due to novelty.

When are speakers fixated?

Finally, in both experiments we completed a time analysis to assess gaze following. We assessed at what time (and the probability of) a participant fixating a target when the target begins their utterance. First, in both experiments we found participants looked at a target speaker earlier in the Control condition versus when the eyes were occluded in the Sunglasses condition. Therefore, it appears viewing the eyes has some beneficial effect on facilitating conversation following. This reflects an abundance of research which describes the use of eyes as a signal (e.g., Ho, Foulsham & Kingstone, 2015), and demonstrates the effect of observing the eyes even in large, dynamic group settings. There were no differences for high and low traits for the time at which participants looked to the speaking target.

We then compared the probability of looks during this plus/minus one second time period. Again, in Experiment 7, there were no differences in the ASD group. However, in both experiments, when the eyes were visible in the Control condition, speaking targets within this time frame were looked at more than in the Sunglasses condition. This indicates

speaking targets were more likely to be fixated within a one second interval of the speech beginning when eyes were not occluded.

In Experiment 8, we did see differences in the probability of fixating a speaking target within the time frame of interest, with a higher probability for ADHD-LT than ADHD-HT. As ADHD is classified as a difficulty in concentration and inattention (American Psychiatric Association, 2013) this could be explained in that participants found it slightly more difficult to locate a current speaker.

Overall, in terms of the theoretical implications of Experiments 7 and 8, we found a strong bias for participants to orient to faces and eyes, and evidence of gaze following of depicted individuals (e.g., End & Gamer, 2019; Flechsenhar et al., 2018; Friesen and Kingstone, 1998). As with all research within this thesis, this has been demonstrated in a more complex and dynamic group setting with fluid, unscripted interactions. In terms of using the eyes as a signalling cue, previously it has been reported that participants in a real interaction use the gaze cues of others (MacDonald & Tatler, 2018). The results from Experiments 7 and 8 provide evidence that information from the eyes is used in multi-party conversation to follow the speaker. This is a unique finding which expands upon gaze behaviours in dyad pairs.

Conclusions

The two experiments gave the opportunity to explore the effect of occluding the eyes in conversation following whilst comparing two population groups of interest. We demonstrate that although overall there are minimal visual attention patterns in overall looks to targets and speakers, there are some diverging results upon deeper analysis. We found ASD-HT participants have decreased attention to the eyes and mouth of speakers than ASD-

LT, even in this larger third-party group setting. Additionally, when looking at the time-based analysis of conversational gaze following, occluding the targets' eyes with sunglasses affected the time course of looks, with participants slower to fixate a speaker upon the utterance beginning. Hence being able to view the targets' eyes facilitated eye-movements to a current speaker. We suggest this was due to the inability to follow the targets' signalling cues (their eyes) which impeded conversation following. Our findings highlight visual attention dissociations between ASD and ADHD in social communication and the impeding effect of occluding the eyes on conversation following.

Experiment 9 - Eye don't understand: sunglasses impede conversation comprehension

This supplementary experiment has been created to focus on the comprehension and information participants extract when attending to the video clips used in prior experiments. In the previous experiments (Experiment 5,6,7 and 8), when presenting the video clips, I included simple questions as an attention check. Although I asked participants these questions, they were not suitable to examine differences in comprehension. Here, I therefore explore how much is understood from the video clips presented (this time without collecting visual attention data).

In the present Experiment, after watching the clips, participants are given a series of questions based on the targets' dialogue. Here, I investigate how and if clips of targets with and without sunglasses occluding their eyes has any effect on comprehension and if the participant's individual ADHD score affect their responses.

Experiment 7 and 8 demonstrated that occluding targets' eyes meant conversation following (that being participants fixating the current speaker upon their utterance beginning), was less likely and later, than when the eyes were visible. In Experiment 8, I found participants in the high trait ADHD were less likely to look to a speaker upon their utterance beginning. Hence here I also explore whether these factors have any effect on conversation comprehension.

Introduction to Experiment 9

The current experiment expands on findings in Experiment 8. In that experiment, we found that occluding the eyes with sunglasses affected conversation following, in that the speaker was fixated less and later upon beginning their turn of talk. This adds support for the use of the eyes as a signalling cue during conversation. In addition, I found that traits of ADHD modulated this, with more looks to the speaker (within this time frame) in the low trait group. This study explores whether these two factors influence what is understood about the conversation. For further information regarding how occluding the eyes may affect attention to conversation see previous chapters.

Conversation comprehension in ADHD

Although not a core diagnostic criterion of ADHD, the DSM-IV (American Psychiatric Association, 2013) suggests there may be an association with language disorder, and in particular in social language skills and pragmatic language deficits (Camarata & Gibson, 1999).

Evidence of different language characteristics are often found in developmental studies. For example, Kim and Kaiser (2000), report children aged six to eight with ADHD perform worse in sentence imitation, word articulation and produce more inappropriate pragmatic behaviours when engaging in conversational interactions. Equally, there is evidence to suggest children with ADHD perform worse in televised story comprehension (Sanchez et al., 1999; Lorch et al., 2006) and overall listening comprehension (McInnes et al., 2003).

Although this thesis does not explore this, it should be noted that there is extensive evidence which explores ADHD and working memory deficits, which arguably could affect language comprehension.

Present study

The present study asks two main questions. First, whether targets wearing sunglasses influences third-party conversation comprehension. I hypothesise that, as there were some observed differences in visual attention in Experiment 8, that these may transpire to affect conversation comprehension. As we use the eyes as a social cue, occluding the eyes may result in a worse performance. Equally, it could be hypothesized that the sunglasses draw attention away from the conversation content, making it more difficult to concentrate on the spoken word.

The second question explored in this study relates to how comprehension performance is linked to ADHD traits. For this study, I will be assessing self-reported traits in the general population. We may expect poorer performance to be linked with higher ADHD self-report scores. However, as this sample is a general population and not classified as explicitly high or low traits, we may see more of a gradual association between the two.

Method

Participants

Participants were 106 (73 female) volunteers, recruited via the online platform Prolific. All participants were native English speakers. Participants were rewarded with a small monetary value, (in line with Prolific's guidelines) for their participation.

Stimuli

The survey was created and presented in Qualtrics. The ADHD questionnaire used was the Adult ADHD Self-Report Scale (ASRS; Kessler et al., 2005), which includes 18 items. As in Experiment 7 and 8, the participants watched 8 video clips of targets having a conversation (2 clips of each of the 4 groups). In half of videos the targets wore sunglasses (Sunglasses Condition, in the other half they did not (Control Condition). For further details of stimuli creation see previous experiments. The participants then answered one open-ended

question per clip based on the content discussed by targets within the clip. Each question had a maximum score of three. An example of the questions included: “Name 3 things the person on the left said.” These questions were piloted for difficulty and prepared with a view to elicit clear correct answers.

Design

The study was a within-subjects design. All participants saw all the clips and took part in all elements of the study. The order of the clips (and accompanying question) was randomised.

Apparatus

Participants completed the online survey using their smartphone, tablet or computer. Before taking part, participants were told they would need a device with sound.

Results

Exclusions

Responses that were not 100% completed were removed from analysis. Qualitative responses from participants which were inappropriate (for example nonsense answers to questions) were removed from analysis completely. Four participants’ responses were excluded from analysis due to inappropriate answers. Therefore 102 respondents were included in analysis.

Qualitative data processing

Previously established correct answers to the questions were then manually compared against the participants’ qualitative responses. Participants were given one point for each correct answer, with a maximum score of 3 per question using a pre-designed marking template. This marking template was piloted for floor and ceiling effects and carefully

constructed to provide a clear guideline for marking. The coding of correct marks was compared with a sample of an additional researchers, to test for inter-rater reliability. A Cronbach's Alpha showed a high and acceptable reliability of α=.98, with a correlation of R= .96, p<.001.

ADHD

The average ADHD score across participants was 5.9 with a range of 0 to 18. For a breakdown of sample size categorised into low (<4), control (4-15) and high (15+), see Table 9.1. Only 4 participants self-reported as high trait ADHD. This low number is expected as roughly 4-5% of the population is thought to be high trait ADHD.

Comprehension

The participants' comprehension scores were converted into percentages to give each participant an overall score. Across all participants and conditions, the average (SD) comprehension score was 74% (13.7) correct. Table 6.13 below shows comprehension scores split by the participants ADHD score. Here, I chose to split the data into these categories, but hereafter (due to the low number of participants in the High group), I chose to look at ADHD as a continuous variable.

ADHD score	*N*	*Mean % correct*	*SD*
Low	34	70.38	15.45
Control	64	75.64	12.71
High	4	80.75	7.93

Table 6.13. The mean correct comprehension scores for participants split by their ADHD score.

Clip condition

I was then interested to see whether occluding the eyes affected comprehension scores. The mean average (SD) comprehension percentage score for the Control and

Sunglasses conditions were 77% (0.17) and 71% (0.17) respectively. A paired samples t-test established that there was a significant difference between these conditions, t (101) =2.95, p=.004. Hence, participant's comprehension scores were higher when watching clips of participants without wearing sunglasses (see Figure 6.6). This demonstrates an impeding effect of occluding the eyes, supporting the results of Experiments 7 and 8.

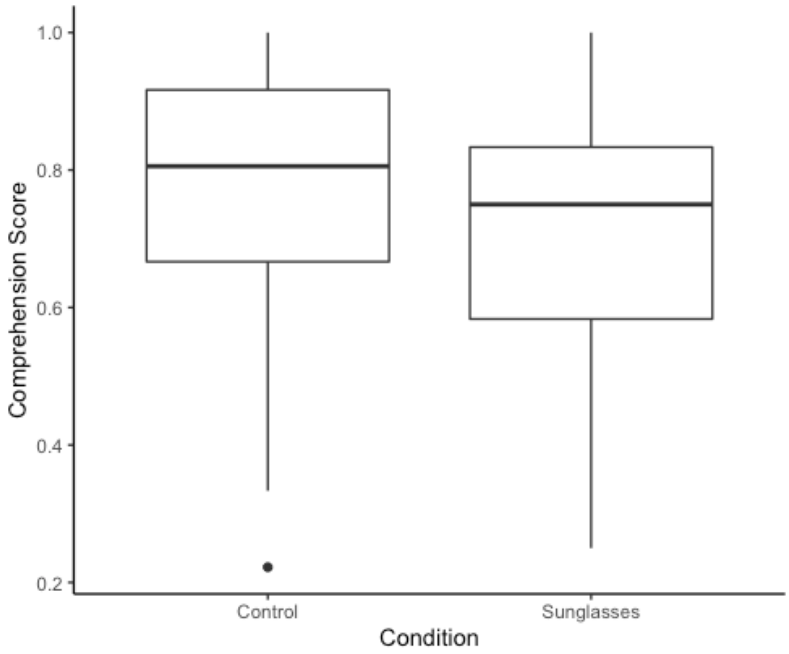

Figure 6.6. Demonstrates participants comprehension scores in each condition. Boxplot to show the participant comprehension scores for each experiment. Boxes show the median and percentiles with whiskers showing the interquartile range and outliers represented at dots beyond.

Score correlations

Overall, there was a significant positive correlation between overall comprehension and ADHD score (regardless of condition), R=.24, p=.015.

When breaking this down further into the two clip conditions, I correlated the two comprehension scores (Sunglasses and Control), with participants' ADHD score. There was a significant and moderate positive correlation between participants' ADHD score and their

Control comprehension score, R=.32, p=.001, while there was no significant correlation between participants' Sunglasses comprehension score and their ADHD score (p>.05). Hence, interestingly, in this instance, when looking at the Control clips, the higher the participants ADHD score, the higher their comprehension score. This is demonstrated in Figure 6.7.

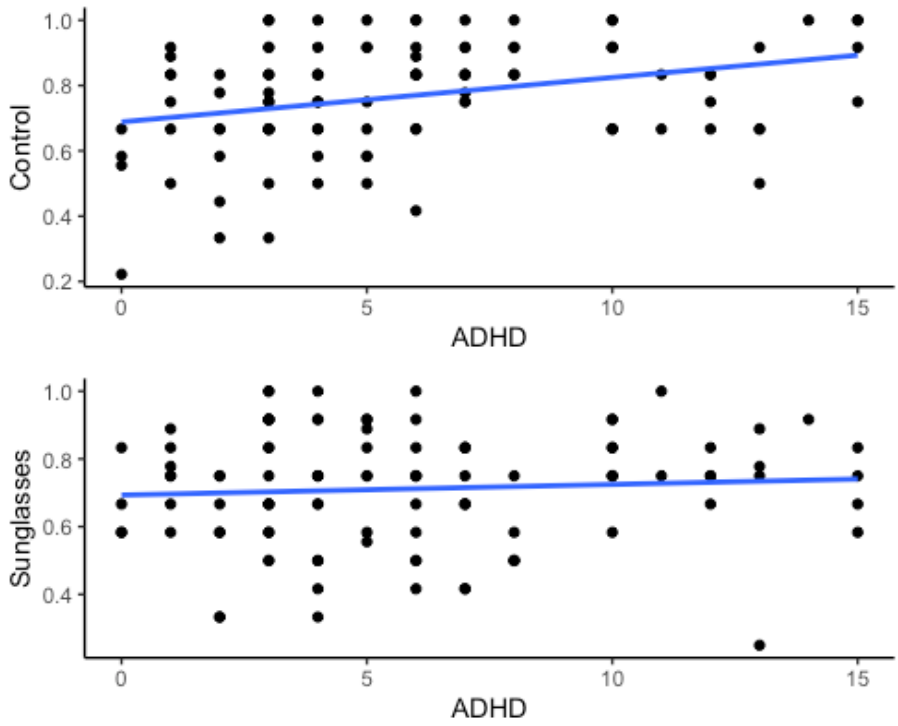

Figure 6.7. Demonstrates the relationship between ADHD scores and Comprehension Scores for the Control (top) and Sunglasses (bottom) conditions.

Liner mixed-effects modelling

I then used a linear mixed effects modelling approach (lmer), which allowed for control of random effects (in this case participant). I used the predictor variables of ADHD and Condition to predict comprehension scores. I used the lme4 package with a Gaussian function, assessing the contribution of each factor with maximum likelihood.

Using a model building approach, I added the continuous variable of ADHD to the intercept only model, with the probability of scoring high on comprehension increasing as

ADHD scores increased (β = .008, ± 0.0034 *SE*). Adding ADHD significantly improved the model ($\chi^2(1) = 6.02, p=.014$). Next, I added Clip Condition to the model, which significantly improved the model fit ($\chi^2(1) = 8.30, p=.004$), with a probability of higher scores in the Control condition when the eyes were visible (β = -.057, ± 0.019 *SE)*. The interaction was also significant, demonstrating that those with a higher ADHD score had better comprehension scores, compared to lower ADHD scores. However, this was found only for the Control condition when the eyes were visible. This effect was not present when the eyes were occluded.

Discussion of Experiment 9

This supplementary online qualitative experiment was created as a means to understand, if, at all, comprehension is affected by participants' ADHD score or the presence of target's eyes. Using the same stimuli in Experiment 7 and 8, I dove deeper to understand whether the visual attention differences seen transpired to conversation comprehension. Findings suggest a surprising positive correlation between ADHD score and comprehension. Interestingly, there was also a significant difference in comprehension scores within the two conditions (Control and Sunglasses) with participants' comprehension of the clips improved when the target's eyes were visible.

ADHD

Based on the previously described literature, which details the impeding effects of ADHD on comprehension, we would expect to see poorer performance associated with higher ADHD scores. In fact, I found the opposite. Overall, higher ADHD scores resulted in better comprehension. One suggestion for this finding could be that the eyes of high trait

ADHD populations move around the scene more, meaning their attention is more distributed. However, this was not what we found in Experiment 8.

A further interesting finding is that there was no significant effect of ADHD scores when participants viewed the Sunglasses condition, with differences only established in the Control condition. This could potentially be explained by the fact the sunglasses make comprehension difficult overall, for both groups. Hence, when conversation is 'normal', that being with visible eyes, the differences in ADHD are more prevalent.

Despite the finding that those with higher ADHD scores had better comprehension performance, I should highlight some key limitations. First, the data collected for ADHD is self-reported. Although the ASRS has high credibility, we must be aware of this as a diagnosed clinical population could produce different results. Second, I did not include only a sub sample of participants who met a certain criterion for ADHD (as we did in Experiment 8). Instead, I collected results from the general population. This therefore meant participants had a range of scores, with the majority (61.5%) falling into the 'Control' category of ADHD traits. Therefore, as expected, only a small proportion of participants fell into the category of high trait. For this reason, the ADHD results should be deliberated with caution.

Despite this finding, ADHD is a complex disorder, with no clear-cut differences in visual attention (as previously described). Therefore, the attention to comprehension is perhaps also as indistinct.

Presence of the eyes

Interestingly, the ability to view the target's eyes did affect comprehension scores, with, as predicted, lower scores when the target's eyes were occluded by sunglasses. This result is more robust given that the conditions were fully between. Hence, all participants saw both conditions giving us scores for each condition which I could directly compare. As suggested, one explanation could be because the sunglasses detract attention from the

conversation, which may have resulted in inattention to the conversation. Arguably, the sunglasses were a novelty, given that the targets are wearing sunglasses inside. However, participants did see multiple clips, which could lead to habituation and were told about the targets wearing sunglasses in the participant instructions. Equally, it could be argued that targets may not appear as interesting to participants when their eyes are occluded. This relates to the 'special processing of the eyes', as is discussed thoroughly in this thesis. However, in Experiment 8 we found no effects of overall visual attention to targets, which perhaps eliminates that suggestion.

Furthermore, another possible explanation is that the targets themselves acted differently when wearing the sunglasses. Although, in the experiments which used these stimuli (7,8 and 9), I made a conscious effort to choose comparable clips, in terms of conversation engagement, it could be that targets made more elaborate movements, were more engaging with their tone of voice, or spoke about more interesting things in the Control clip. There are several ways future studies could target this problem. First, an option could be to use the Control clips and superimpose sunglasses onto the video clips. This would ensure the exact same images are used, with the visibility of the eyes being the only manipulation. The clips could also be staged by actors (with and without sunglasses), but this may remove the 'natural', real-world element, which is a benefit of these experiments. Lastly an option could be to repeat the experiment, but this time use the audio alone. It would be expected that although overall comprehension may be lower (without the visual counterpart), I should not find differences between the two conditions. This would support the finding that being able to view the eyes of others facilitates conversation understanding.

In Experiment 8, we too found differences in conditions when occluding the eyes, in that participants' gaze following was slower and less around changes in speakership. Hence

results in the present experiment add increased evidence for a maladaptive effect of occluding the eyes when observing conversation.

Conclusion

Experiment 9 explored the effect of removing the eyes as a signalling cue when comprehending third-party group conversation. Overall better comprehension scores were positively correlated with being able to see the eyes and oddly, higher ADHD scores. However, this result should be treated with caution given the small sample of high trait individuals in the general population. This experiment adds additional support for the use of eyes in conversation following and helps us to understand what information is extracted when viewing group interactions.

Chapter Summary

This chapter has discussed three experiments which explore attention during conversation with regard to how traits of ASD and ADHD modify behaviours. Overall, I found decreased attention to social areas in the high trait ASD group, which offers a confirmatory result when moving to larger groups (in line with dyad interactions). The effect of ADHD traits is not as straight forward, with contradictory results. This confirms how complex the disorder is. In terms of the effect of sunglasses occluding targets eyes, I found supporting evidence that this disrupts attention to conversation. Not only were participants slower and less likely to look at a target speaker upon utterances beginning (Experiment 7 and 8), but occluding the eyes also affected conversation comprehension (Experiment 9). Altogether, this supports the importance of using the eyes as a signalling cue during group conversation.

General Discussion

This thesis has explored how attention is allocated to people in complex settings. More specifically, Experiments 1-3 involved judging images of mobile eye tracking data, whilst Experiments 4-9 explored how we visually attend to real-world, dynamic conversations. Here, I have investigated what attracts visual attention to interlocutors within interactions with various manipulations.

As reviewed in Chapter 1, to date there are limited explorations of visual attention to conversation in more 'real-world', dynamic and complex settings. This is surprising given that the majority of our day-to-day conversations do not involve simple turn-taking with one other person. To date, most of the research which explores visual social attention involves dyad pairs engaging in conversation. Previous research of this kind has established that social elements of the scene (e.g. a person's face) attract more visual attention (e.g. Pascalis et al., 1998; Theeuwes & Van der Stigchel, 2006; Kendon, 1967; End & Gamer, 2017), that interaction (Skripkauskaite, Mihai & Koldewyn, 2021) and speakership modulate this (e.g. Argyle & Ingham, 1972; Hirvenkari et al., 2013), and that there are systematic timing patterns in the time course of gaze to a speaker (Ho, Foulsham & Kingstone, 2015). However, the research to date is often scripted, simple and unrealistic (Risko et al., 2016). To tackle this, this thesis has investigated visual attention to group conversation using stimuli which reflects real-world scenarios.

A key question which this research has addressed is: where do we look when someone is speaking? Previous literature has suggested we look to a person when they are talking (e.g. Hirvenkari et al., 2013) as a social cue to show we are listening or perhaps to better understand the conversation. Using a range of techniques, I have investigated: how this differs when moving to larger more complex group settings, whether there are any differences in live versus third-party viewing, which audiovisual cues facilitate this, whether

occluding the eyes has any affect, and finally whether traits of ADHD and ASD modify viewing behaviours.

As an additional line of research, I also explored how subjective the manual coding of MET data can be (as this is a technique often used in visual social attention research). A total of nine experiments are reported in this thesis which address these questions.

Overall findings

First, I explored the methodological concerns when conducting mobile eye tracking (MET) data analysis. MET is an attractive technique, especially for research with a social element and for the overall research questions of this thesis. Arguably MET can allow for greater ecological validity, but it does not come without challenges. Not only can manual coding be laborious, but the coder themselves may have biases. Using three experiments it was demonstrated that assumptions about other people's gaze (ToM) make coding MET cursors a subjective experience. In Experiment 1, we confirmed that describing the cursor as 'biodata' resulted in participants adopting another's perspective and imposing their own ToM biases onto the cursor. The results demonstrated that participants were more likely to decide a MET cursor was 'on' a face than an object, despite being located at the same distance away from the target. This result was abolished and minimised in Experiment 2 and 3, respectively, where the biodata aspect was removed. In other words, when knowing a cursor is from an eye-tracker, manual coding is a more subjective experience (when coding social scenes). This was something I personally experienced upon manually coding MET data and my results in Experiments 1, 2 and 3 confirmed my suspicions. From these research findings, I strongly suggest there should be strict coding rules decided prior to analysis and followed by all coders. Additionally, an awareness of this bias may aid objectivity of the analysis. This research had implications for Experiment 4 which used MET, as well as other researcher's past and future studies which use manual coding of MET data.

Chapter 3 included Experiment 4, which investigated live, interactive looking behaviours when directly comparing dyad and triads, with and without direct eye contact. This experiment is classed as exploratory due to data concerns. As highlighted in Chapter 2, MET data can be extremely temperamental and subjective. I remain cautious about these findings due to some problems with data quality (a large number of exclusions and data variability). Despite this, there were some differences in group size, with more social attention in a dyad than a triad, which could be explained in terms of social loafing. Additionally, participants did show a general pattern of looking more to others while listening and averting their gaze while speaking. Due to the problems in MET, the remainder of the eye tracking experiments (5-8) used a desk mounted static eye-tracker with improved accuracy.

That being said, Experiment 5 offered a unique approach to directly compare live and third-party visual attention in larger groups. The groups were made up of six individuals and (with ecological validity in mind) depicted very natural and fluid conversation. Promisingly, the live and third-party looking behaviours in conversation were very similar. To my knowledge, this experiment is the first to directly compare the two in the same situation, which has great significance for the social attention field. I then further explored manipulations of third-party clips, with various elements which could affect viewing including audiovisual manipulations (Experiment 5) and the spatial location of the targets (Experiment 6). In Experiment 5, I used four video clip manipulations to examine how participants follow the conversation whilst signalling cues were manipulated. These were a control condition, a sound off condition, a freeze-framed condition, and a blank screen condition, with the audio continuing in the latter two. I found participants were able to follow conversation with one modality (audio or visual). However, gaze to a speaker was strongest when both modalities were available. Interestingly there was evidence of gaze following in

the freeze-frame condition, despite no additional visual information being gained by fixating targets. To explore this further, I created Experiment 6. The aims of this study were to first investigate whether this effect could be replicated and second to explore whether manipulating the expected location of the targets affected gaze following. Overall, there were minimal differences in the freeze-frame conditions, even with a spatial change of target location. These results confirmed findings in Experiment 5. Suggestions as to why this occurs can either be due to the social elements continuing to attract attention or perhaps viewing a target helps with conversation understanding. These studies explored key elements of how different spoken and physical movements affect following in turn-taking conversation, but in a larger more dynamic composition. In other words, which cues facilitate gaze following during group conversation.

Continuing an investigation into signalling cues, Experiments 7, 8 and 9 used third-party observation to explore the effect of occluding the eyes. In half of the clips shown to participants within these experiments, targets wore sunglasses. This removed the ability to use the eyes as a signalling cue in gaze following. Overall, occluding the eyes with sunglasses did impede conversation following and understanding. In Experiments 7 and 8, participants were faster and more likely to fixate a speaker upon their utterance beginning when the eyes were available. This demonstrates how participants use the eyes of others to recognise turn-taking, even in third-party large group conversations. In Experiment 9, participants' comprehension of the same clips was also negatively impacted when targets wore sunglasses.

In addition to occluding the eyes, the final three experiments also explored the effects of clinical traits on visual attention to groups. In Experiment 7, I added the additional manipulation of high and low ASD traits. In line with previous research, high trait ASD participants showed decreased attention to the eyes and mouths of targets than low trait

participants. This result was hence confirmed in larger, more complex groups. Furthermore, Experiment 8 and 9 assessed the effect of ADHD. There were minimal and contradictory differences of ADHD, demonstrating the complexity of the disorder presentation.

Overall, these findings helped us to paint a clearer picture to understand how various aspects affect visual attention to social conversation in dynamic group settings.

Theoretical significance

Problems with mobile eye tracking

A concerning topic highlighted within this thesis was the difficulties of using mobile eye tracking data to explore my research questions. Experiment 4 demonstrated the problems of collecting this data whereas Experiments 1-3 investigated differences in subjectivity when coding the data. Experiments 4 unfortunately had poor data quality which meant it was particularly difficult to perform inferential statistical analysis and draw conclusions from this dataset. This thesis presented a very honest account of the poor data and, despite substantial efforts to control for these expected problems, the data was of unacceptable quality. With this in mind, questions arise as to the accuracy of previous research that uses this method. Further problems associated with MET were highlighted in Experiments 1-3, where I experimentally demonstrated the coding subjectivity which I hypothesized may be occurring. Here, people were more likely to code a cursor as on a face than an object, despite being of equal distance away from the target. Conclusions from these experiments were that participants were attaching theory of mind to the cursor, when they believed it reflected biodata. In other words, knowing this cursor was a gaze location from another human observer meant we are more likely to assume they are looking at a face than an object. Hence, our own biases are interacting with our ability to objectively code a cursor's location. Given the overwhelming literature which suggests we are compelled to look to faces (as cited in Chapter 1), this may not seem too surprising. Additionally, in the eye tracking video experiments to follow

(Experiments 5-8), we also see increased visual attention to faces. However, if the reason for results in Experiment 1 were due to the coder's increase in attention to faces, we would have seen similar results in Experiments 2 and 3, but this was not the case. Instead, as stated, a ToM element is involved where we impose our own biases onto others gaze locations.

Questions then surface as to whether the costs of mobile eye tracking outweigh the benefits. Of course, there will be some circumstances where a mobile device will be more appropriate and effective. For example, perhaps in situations where locomotion is necessary. In Experiment 5, I established that in group conversation live interlocutors and third-party viewers viewed targets similarly. This supports prior and future work which generalises observation of video to real-world situations. Therefore, arguably in this context, the costs of using MET as a method of data collection outweigh the benefits.

Looking to a speaker (in video and real life)

Looking to a person who is speaking is a vital part of successful communication. As readily discussed in this thesis, we tend to look to social aspects of our live environment and also in a pre-recorded scene (Foulsham, Walker and Kingstone, 2011). Speaking has also been shown to modulate this, with an increase in visual attention to a current speaker (e.g. Hautala et al., 2016). This thesis first investigated whether this was true for larger, more dynamic groups and second how this differed in both lab and real-world.

Although it was true that participants looked to a speaker, in the presented experiments, the percentage of looks to speakers was smaller than anticipated. For example, in the Control condition in Experiment 5, although 98% of looks were on targets, only 59% of fixations were on a current speaker. This perhaps suggests that as more speakers are added to the conversation, it may make conversation harder to follow. Unfortunately, whether looks to the speaker were modulated by group size could not be explicitly tested given the data quality of Experiment 4. However, future work could explore the effect of group size using

third-party observers. Equally, looks to non-speakers could be caused by participants relying on the other targets to help comprehend the conversation. It may be that the other targets reactions to the speaker were more visually interesting. This is a plausible explanation given the choice of stimuli. As stated, I chose groups who were naturally formed, thus eliminating awkward and unnatural dialect. This further adds to the literature given these results were established from stimuli of this nature.

An additional analysis, only touched upon in this thesis, but which has considerable merit, would be to assess the target's own eye and head movements in Experiments 5-8. Here, a further avenue of research could explore how the targets' eye and head movements (hence their own visual attention) affects third-party looks to a speaker.

Furthermore, in Experiment 5 I used a unique method to compare a live scenario with people watching a recording. Here I found similar looking behaviours to targets in both settings, demonstrating the strength of looking to targets (and a similar proportion of looks to speakers) in both situations. Questions arise as to why this is observed. In a live social setting, we look to a speaker to gain information but also to signal that we are listening. If you were to imagine you were engaged in a group conversation and you directed a large proportion of your visual attention to a non-speaker, this would appear very abnormal. Furthermore, with social rules informing us that it is rude not to look at a speaker, these findings in a live conversation are not surprising. However, if we think about why we follow conversation in video, the need to provide a social cue to our interlocutors that we are listening is removed. Hence, there must be an additional benefit to visually following the conversation. The results of Experiment 6 may suggest this helps us to better understand the conversation or equally this may be just a habit. To better answer this question, future research could explore how the task at hand affects visual attention to groups. Overall, this

thesis has provided knowledge of gaze to speakers in larger, more dynamic and real-world group settings.

Guiding looks to a speaker

Although it is well established that we look to a speaker, the cues which guide us to the gaze location are less certain. This thesis explored which cues facilitate looking at a current speaker. As discussed, a speaker may receive visual attention for the physical or auditory features they express (for example their movement or speech). Experiment 5 explored whether a visual or auditory counterpart is more effective in gaze following. Overall, participants followed gaze most effectively when both modalities were present (in line with past research), followed by the visual only and the auditory only conditions respectively. With this in mind, Experiments 7-9 investigated how a visual cue may be facilitating gaze following, by removing the ability to observe (arguably) the most influential cue - the eyes. As was discussed in Chapter 1 'The Cooperative Eye Hypothesis' (Tomasello et al., 2007), the eyes appear to be vital to successful communication. By removing this visual cue, in Experiments 7-9 I found an impeding effect. Whether there were also any differences reflected in a live situation was not proven (due to data issues in Experiment 4).

How this affected timing of looks to a current speaker was also an integral part of this research. In the past, there is inconsistent evidence of whether there are anticipatory looks to a speaker (see Introduction of Experiment 5). In the experiments in this thesis, I did not find anticipatory effects, with looks to speakers after roughly a 500ms delay of utterance beginning (similar results in Experiments 5, 7 and 8). Given the complexity of the group conversation (and moving away from scripted dyad encounters), this isn't too surprising. However, this thesis did find a number of factors that influenced the timing of looks to speakers. In Experiment 5, it was found the visual counterpart (as opposed to auditory) helped guide participants to a current speaker and in Experiments 7 and 8, being able to

observe the eyes also acted as a facilitator. All of this helped to comprehend how conversation is followed in real-world scenes.

Reflection

From the above general discussion points, it is apparent there is an abundance of knowledge gained from this work. In particular, there are specific findings within Experiments 1-3, 5 and 7-8 which offer significant learning opportunities and advance knowledge of the field.

Upon reflection, Experiments 1-3 not only teach us how a simple decision can be subjective in a social context, but also how theory of mind interplays with this decision. Within the three experiments, I asked participants a simple question: "Is the cursor on the target?". Despite identical pictures being used, I demonstrated how the social context and ability to adopt visual perspective modulated decisions. In Experiment 1, where a cursor was described as representing a person's eye gaze, participants were more likely to decide a cursor was on a face than an object (even when at the same distance, ($\chi^2_{(4)} = 206.12$, $p<.001$)). This effect was eradicated (($\chi^2_{(1)} = 3.69$, $p=.055$) when we changed the story, by describing the cursor as randomly generated by a computer and minimised ($\chi^2_{(1)} =15.74$ $p<.001$), but still evident, when the cursor was described as a computer actively seeking the target. As is seen clearly in Figure 2.7, Experiment 1 yielded a considerably stronger face bias than the other two experiments. This suggests participants are attaching what they implicitly know about ToM and perspective taking to the gaze location of another human. This in turn affects the interpretation of the stimulus. Collectively, these experiments demonstrate how ToM affects even the simplest of decisions – is a cursor on or off a target. This in turn has significant impact for the field of ToM as it demonstrates how, in certain

contexts, a motionless and seemingly meaningless cursor point can hold representations of other people's gaze and attributions of the mind, which in turn, influence behaviour.

Experiment 5 is unique in that it is the first to directly compare eye movements in a live situation to those watching the same interaction pre-recorded at a later stage. Here, I found that the percentage of time looking to a speaking target was remarkably similar (59% in the live interaction and 51% in the third-party viewing). When also comparing the timing of looks to each target, again there was a high agreement with 60% of looks to the same target at the same time with a mean kappa of .79. This finding in particular is of increased importance given the nature of lab based social visual attention research, which is often criticised for its ecological validity. From this experiment, promisingly we learnt that behaviour in the two situations is comparable. This advances our field two-fold. First, by using a unique design to explore the effect of social presence in social interactions. Second, by insinuating visual attention is similar in the two settings, this work provides evidence for using such methods to explore attention in dynamic interactions. For example, using third-party viewing, which is more cost effective in data collection, offers a good proxy for the assessment of live gaze behaviour.

A further learning point within this thesis is the finding that occluding the eyes inhibits conversation following. Experiments 7 and 8, demonstrated how participants were slower (Exp 7= F (1,39) =5.22, p<.05, Exp 8 = F (1,37) =13.82, p<.05) and less likely to fixate a speaker within 1 second of the speaker's utterance beginning when the targets within the scene were wearing sunglasses than when not. In prior research, it is evident how we use the eyes of others to facilitate cooperation and this present finding is noteworthy given it extends this to a complex and dynamic group interaction. This in turn advances our knowledge of how the eyes are 'special' and vital to fluid social communication.

Overall, the research within this thesis offers a wide-ranging account of social gaze in dynamic contexts, all of which advances our understanding of the field of social attention.

Future directions and implications

Implications of this research in social settings are vast. This thesis provides an understanding of how gaze operates in a social interaction expanding on past research, pushing the methods of previous work to include multiple interlocutors, in a more real-world setting. Additionally, this thesis uses a combination of both live and third party viewing which can be compared in their findings, allowing a reflection of how lab studies can be compared to real behaviours. I used both static and mobile eye tracking to infer the results adding further depth to the main research question. In doing so, this research expands upon more simple dyadic interactions and explores what we know about visual attention to larger realistic groups. A key finding of interest is the comforting similarities of third-party and live looking behaviours to these larger groups. This result can be used as support for future work using third-party methods, which (as is found within this thesis), are more accurate and generalisable.

Successful gaze behaviours in social situations are vital to healthy development, with gaze commonly used as a learning mechanism. Gaze, although often subconscious and automatic, is incredibly important to facilitate cooperation. Gaze is used to guide the flow of effective interactions. Hence, the work presented within this thesis can be applied both to human behaviours as well as robotics and machine learning with virtual agents.

For example, findings can be used to assist computer scientists who are placing increased importance on robots who can interact with humans. Fadda et al. (2020) describe how scientists are now attempting to create social robots who can engage and interact with humans in the real world. They explain how psychology is being used to explore gaze

following, which in turn can be applied to robots. Hence, the findings presented in this research have particular relevance to robotics and AI.

Furthermore, if we fully understand gaze behaviours in a 'healthy' population, we can shed light on disorders which present atypical visual attention. This thesis research may help us understand any differences within ASD and ADHD populations.

In Experiment 4, I explored the effects of group size on gaze behaviour. This has implications for other areas of applied research. For example, one may be interested in knowledge of dominance and leadership abilities, which could be analysed in terms of job interviews (Maran et al., 2020) and other team tasks. Here, perhaps gaze could be used to better understand leadership in groups or even predict job success if working in a larger team.

Given the recent pandemic, where our social interactions have changed dramatically, future directions could explore visual attention to social interaction in online settings. For example, would the effects of group size, eye presence and the effects of audiovisual information reported in this thesis be present in a Zoom set up? If conversation following was also impeded by occluding the eyes in video calls, it would demonstrate the importance of eyes even in online virtual environments. Furthermore, in Experiment 5 I established how conversation following occurred in blank and freeze-frame conditions whilst audio continued. Here, participants visually attended to the current speaker more so in the freeze-frame condition, where the targets' faces (static images) were present. This could perhaps influence the design of Zoom and other similar software, whereby a static image of a person could be presented instead of an empty box when videos are turned off.

Additionally, given the particular focus of visual attention on facial features during conversation, questions arise as to how wearing masks, which conceal the mouth area, affect fixations to the face. Would we find that the eyes are even more 'special' given the occlusion

of the mouth? This is something which would be a fantastic and topical next step for the research presented in this thesis.

Closing remarks

Overall, this thesis offers a wide-ranging, comprehensive account of how visual attention is directed during social conversation in both third-party and live settings. By exploring a range of factors which can affect gaze, this thesis provides an abundance of research which strengthens our understanding of social visual attention in real-world settings.

References

Allison, C., Auyeung, B., & Baron-Cohen, S. (2012). Autism Spectrum Quotient. *Journal of the American Academy of Child and Adolescent Psychiatry, 51*(2). 202-12.

Apperly, I. A., & Butterfill, S. A. (2009). Do humans have two systems to track beliefs and belief-like states? Psychological review, 116(4), 953.

Altmann, G.T.M. (2004). Language-mediated eye movements in the absence of a visual world: the 'blank screen paradigm'. Cognition 93, 79–87.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders: DSM-5*, 5th ed. Washington (D.C.).

Argyle, M., & Ingham, R. (1972). Gaze, Mutual Gaze, and Proximity. *Semiotica*, *6*(1), 32–49.

Baron-Cohen, S. (1995). *Mindblindness*, MIT Press: Cambridge, MA.

Baron-Cohen, S. (1995). The eye direction detector (EDD) and the shared attention mechanism (SAM): Two cases for evolutionary psychology. In C. Moore, & P. J. Dunham (Eds.), Joint attention: Its origins and role in development (pp. 41–59). Hillsdale, NJ: Lawrence Erlbaum.

Bayliss, A. P., Di Pellegrino, G., & Tipper, S. P. (2004). Orienting of attention via observed eye gaze is head-centred. *Cognition*, *94*(1).

Beebe, S. A. (1974). Eye contact: A nonverbal determinant of speaker credibility. *Communication Education*, *23*(1), 21-25.

Benson, V., & Fletcher-Watson, S. (2011). Eye movements in autism. *Oxford Handbook of Eye Movements*, 709-730.

Birmingham, E., Bischof, W. F., & Kingstone, A. (2007). Why do we look at people's eyes?. *Journal of Eye Movement Research*, *1*(1).

Birmingham, E., Bischof, W. F., & Kingstone, A. (2008). Social attention and real-world scenes: The roles of action, competition and social content. *The Quarterly Journal of Experimental Psychology*, *61*(7), 986-998.

Birmingham E. Kingstone A. (2009). Saliency does not account for fixations to eyes within social scenes. *Vision Research*, *49*, 2992–3000.

Boccignone, G., Cuculo, V., D'Amelio, A., Grossi, G., & Lanzarotti, R. (2020). On gaze deployment to audio-visual cues of social interactions. *IEEE Access*, *8*, 161630-161654.

Bolis, D., Balsters, J., Wenderoth, N., Becchio, C., & Schilbach, L. (2017). Beyond autism: introducing the dialectical misattunement hypothesis and a Bayesian account of intersubjectivity. *Psychopathology*, *50*(6), 355-372.

Boucher, J. D., Pattacini, U., Lelong, A., Bailly, G., Elisei, F., Fagel, S., ... & Ventre-Dominey, J. (2012). I reach faster when I see you look: gaze effects in human–human and human–robot face-to-face cooperation. *Frontiers in neurorobotics*, *6*, 3.

Broadbent, D. E. (1956). Successive responses to simultaneous stimuli. *Quarterly Journal of Experimental Psychology*, *8*(4), 145-152.

Brône, G., Oben, B., & Goedemé, T. (2011). Towards a more effective method for analyzing mobile eye-tracking data: Integrating gaze data with object recognition algorithms. *PETMEI'11 - Proceedings of the 1st International Workshop on Pervasive Eye Tracking and Mobile Eye-Based Interaction*, 53–56.

Brunswick, E. (1947). *Systematic and Representative Deisgn of Psychological Experiments*. University of California Press.

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

Buchan, J. N., Paré, M., & Munhall, K. G. (2008). The effect of varying talker identity and listening conditions on gaze behavior during audiovisual speech perception. *Brain*

Research, *1242*, 162–171.
Bundesen, C. (1990). A theory of visual attention. *Psychological Review*, *97*(4), 523–547.
Buschman, T. J., & Miller, E. K. (2007). Top-down versus bottom-up control of attention in the prefrontal and posterior parietal cortices. *Science*, *315*(5820), 1860–1864.
Bushnell, I. W. R. (2001). Mother's Face Recognition in Newborn Infants: Learning and Memory. *Infant and Child Development*, *10*(1–2), 67–74.
Cañigueral, R., Ward, J. A., & Hamilton, A. F. D. C. (2021). Effects of being watched on eye gaze and facial displays of typical and autistic individuals during conversation. *Autism*, *25*(1), 210-226.
Camarata, S. M., & Gibson, T. (1999). Pragmatic language deficits in attention-deficit hyperactivity disorder (ADHD). *Mental retardation and developmental disabilities research reviews*, *5*(3), 207-214.
Campbell, R., Heywood, C. A., Cowey, A., Regard, M., & Landis, T. (1990). Sensitivity to eye gaze in prosopagnosic patients and monkeys with superior temporal sulcus ablation. *Neuropsychologia*, *28*(11), 1123–1142.
Carpenter, M., Nagell, K., & Tomasello, M. (1988). Social Cognition, Joint Attention, and Communicative Competence From 9 to 15 Months of Age - PubMed. *Monographs of the Society for Research in Child Development*, *4*(63), 1–143.
Castellanos, F. X., Marvasti, F. F., Ducharme, J. L., Walter, J. M., Israel, M. E., Krain, A., ... & Hommer, D. W. (2000). Executive function oculomotor tasks in girls with ADHD. *Journal of the American Academy of Child & Adolescent Psychiatry*, *39*(5), 644-650.
CDC. (2020). *Symptoms and Diagnosis of ADHD*. Centers for Disease Control and Prevention. https://www.cdc.gov/ncbddd/adhd/diagnosis.html
Cerf, M., Harel, J., Einhäuser, W., & Koch, C. (2008). Predicting human gaze using low-level saliency combined with face detection. *Advances in neural information processing systems*, *20*, 1-7.
Chawarska, K., & Shic, F. (2009). Looking but not seeing: Atypical visual scanning and recognition of faces in 2 and 4-Year-old children with Autism spectrum disorder. *Journal of Autism and Developmental Disorders*, *39*(12), 1663–1672.
Cherry, E. C. (1953). Some experiments on the perception of speech with one and with two ears. J. Acoust. Soc. Amer. 25, 975–979.
Chita-Tegmark, M. (2016). Social attention in ASD: A review and meta-analysis of eye-tracking studies. *Research in developmental disabilities*, *48*, 79-93.
Chua, H. F., Boland, J. E., & Nisbett, R. E. (2005). Cultural variation in eye movements during scene perception. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(35), 12629–12633.
Clark, H. H., & Brennan, S. E. (1991). *Grounding in communication*. In L. B. Resnick, J. M. Levine, & S. D. Teasley (Eds.), American Psychological Association. *Perspectives on socially shared cognition*, 127–149.
Clarke, A, D, F., Mahon, A., Irvine A., & Hunt, A, R. (2017). "People are unable to recognize or report on their own eye movements.", *The Quarterly Journal of Experimental Psychology, 70* (11), 2251-2270.
Cole, G. G., Atkinson, M., Le, A. T. D., & Smith, D. T. (2016). Do humans spontaneously take the perspective of others? *Acta Psychologica*, *164*, 165–168.
Cole, G. G., & Millett, A. C. (2019). The closing of the theory of mind: A critique of perspective-taking. *Psychon Bull Rev 26*, 1787–1802.
Collis, G. M., & Schaffer, H. R. (1975). Synchronization of visual attention in mother-infant pairs. *Journal of Child Psychology and Psychiatry*, *16*(4), 315-320.
Corkum, V., & Moore, C. (1998). The origins of joint visual attention in

infants. *Developmental psychology*, *34*(1), 28.
Crosby, J. R., Monin, B., & Richardson, D. (2008). Where do we look during potentially offensive behavior?. *Psychological Science*, *19*(3), 226-228.
Dalton, K. M., Nacewicz, B. M., Johnstone, T., Schaefer, H. S., Gernsbacher, M. A., Goldsmith, H. H., Alexander, A. L., & Davidson, R. J. (2005). Gaze fixation and the neural circuitry of face processing in autism. *Nature Neuroscience*, *8*(4), 519–526.
Davidson, G. L., & Clayton, N. S. (2016). New perspectives in gaze sensitivity research. *Learning & Behavior*, *44*(1), 9-17.
Dawson, J., & Foulsham, T. (2021). Your turn to speak? Audiovisual social attention in the lab and in the wild. *Visual Cognition*, 1-19.
De Silva, S., Dayarathna, S., Ariyarathne, G., Meedeniya, D., Jayarathna, S., Michalek, A. M., & Jayawardena, G. (2019). A rule-based system for ADHD identification using eye movement data. In *2019 Moratuwa Engineering Research Conference (MERCon),* IEEE. 538-543.
Deans, P., O'Laughlin, L., Brubaker, B., Gay, N., & Krug, D. (2010). Use of eye movement tracking in the differential diagnosis of attention deficit hyperactivity disorder (ADHD) and reading disability. *Psychology*, *1*(04), 238.
Desimone, R., & Duncan, J. (1995). Neural Mechanisms of Selective Visual Attention. *Annual Review of Neuroscience*, *18*(1), 193–222.
Dink, J. W., & Ferguson, B. (2015). eyetrackingR: An R Library for Eye-tracking Data Analysis. Retrieved from http://www.eyetrackingr.com.
Doherty-Sneddon, G., Anderson, A., O'malley, C., Langton, S., Garrod, S., & Bruce, V. (1997). Face-to-face and video-mediated communication: A comparison of dialogue structure and task performance. *Journal of experimental psychology: applied*, *3*(2), 105.
Dorr, M., Martinetz, T., Gegenfurtner, K. R., & Barth, E. (2010). Variability of eye movements when viewing dynamic natural scenes. *Journal of Vision*, *10*(10).
Driver, J., & Baylis, G. C. (1989). Movement and Visual Attention: The Spotlight Metaphor Breaks Down. *Journal of Experimental Psychology: Human Perception and Performance*, *15*(3), 448–456.
Driver IV, J., Davis, G., Ricciardelli, P., Kidd, P., Maxwell, E., & Baron-Cohen, S. (1999). Gaze perception triggers reflexive visuospatial orienting. *Visual cognition*, *6*(5), 509-540.
Duchowski, A. T. (2017). Eye tracking methodology: Theory and practice: Third edition. In *Eye Tracking Methodology: Theory and Practice: Third Edition*. Springer International Publishing.
Ekman, P., Friesen, W. V., & Ellsworth, P. (2013). *Emotion in the human face: Guidelines for research and an integration of findings,* (11). Elsevier.
Ellsworth, P. C., Carlsmith, J. M., & Henson, A. (1972). The stare as a stimulus to flight in human subjects: A series of field experiments. *Journal of Personality and Social Psychology, 21*(3), 302–311.
Emery, N. J. (2000). The eyes have it: the neuroethology, function and evolution of social gaze. *Neuroscience & Biobehavioral Reviews, 24 (6),* 581-604.
End, A., & Gamer, M. (2017). Preferential processing of social features and their interplay with physical saliency in complex naturalistic scenes. *Frontiers in psychology*, *8*, 418.
End, A., & Gamer, M. (2019). Task instructions can accelerate the early preference for social features in naturalistic scenes. *Royal Society open science*, *6*(3), 180596.
Fadda, R., Congiu, S., Doneddu, G., & Striano, T. (2020). Inspiring Robots: Developmental trajectories of gaze following in humans. *Rivista internazionale di Filosofia e Psicologia*, *11*(2), 211-222.
Falck-Ytter, T., von Hofsten, C., Gillberg, C., & Fernell, E. (2013). Visualization and

analysis of eye movement data from children with typical and atypical development. *Journal of autism and developmental disorders*, *43*(10), 2249-2258.
Farroni, T., Csibra, G., Simion, F., & Johnson, M. H. (2002). Eye contact detection in humans from birth. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(14), 9602–9605.
Ferreira, F., Apel, J., & Henderson, J. M. (2008). Taking a new look at looking at nothing. *Trends in cognitive sciences*, *12*(11), 405-410.
Fink, B., & Penton-Voak, I. (2002). Evolutionary psychology of facial attractiveness. Current Directions in Psychological Science, 11(5), 154-158.
Flechsenhar, A., Rösler, L., & Gamer, M. (2018). Attentional selection of social features persists despite restricted bottom-up information and affects temporal viewing dynamics. *Scientific reports*, *8*(1), 1-10.
Foulsham, T., Cheng, J. T., Tracy, J. L., Henrich, J., & Kingstone, A. (2010). Gaze allocation in a dynamic situation: Effects of social status and speaking. *Cognition*, *117*(3), 319–331.
Foulsham, T., & Kingstone, A. (2013). Where have eye been? Observers can recognise their own fixations. *Perception*, *42* (10), 1085-1089.
Foulsham, T., & Kingstone, A. (2017). Are fixations in static natural scenes a useful predictor of attention in the real world?. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *71*(2), 172.
Foulsham, T., & Lock, M. (2014). How the Eyes Tell Lies: Social Gaze During a Preference Task. *Cognitive Science, 39*, 1704-1726.
Foulsham, T., & Sanderson, L. A. (2013). Look who's talking? Sound changes gaze behaviour in a dynamic social scene. *Visual Cognition*, *21*(7), 922–944.
Foulsham, T., Walker, E., & Kingstone, A. (2011). The where, what and when of gaze allocation in the lab and the natural environment. *Vision Research*, *51* (17), 1920–1931.
Frank, M. C., Vul, E., & Johnson, S. P. (2009). Development of infants' attention to faces during the first year. *Cognition*, *110*(2), 160–170.
Freeth, M., & Bugembe, P. (2019). Social partner gaze direction and conversational phase; factors affecting social attention during face-to-face conversations in autistic adults?. *Autism*, *23*(2), 503-513.
Freeth, M., Foulsham, T., & Kingstone, A. (2013) What Affects Social Attention? Social Presence, Eye Contact and Autistic Traits. *PLoS ONE 8(1)*.
Frischen, A., Bayliss, A. P., & Tipper, S. P. (2007). Gaze cueing of attention: Visual attention, social cognition, and individual differences. *Psychological Bulletin, 133 (4),* 694–724.
Friesen, C. K., & Kingstone, A. (1998). The eyes have it! Reflexive orienting is triggered by nonpredictive gaze. *Psychonomic Bulletin and Review*, *5*(3), 490–495.
García, J. O. P., Ehlers, K. R., & Tylén, K. (2017). Bodily constraints contributing to multimodal referentiality in humans: The contribution of a de-pigmented sclera to proto-declaratives. *Language & Communication*, *54*, 73-81.
Gobel, M. S., Kim, H. S., & Richardson, D. C. (2015). The dual function of social gaze. Cognition, 136(0), 359-364. doi:http://dx.doi.org/10.1016/j.cognition.2014.11.040.
Goffman, E. (1966). *Behavior in public places. Notes on the social organization of gatherings*. The Free Press.
Gregory, N. J., López, B., Graham, G., Marshman, P., Bate, S., & Kargas, N. (2015). Reduced Gaze Following and Attention to Heads when Viewing a "Live" Social Scene. PLoS One, 10(4), e0121792. doi:10.1371/journal.pone.0121792.
Grossman, E. D., & Blake, R. (2001). Brain activity evoked by inverted and imagined

biological motion. *Vision Research, 41* (10-11), 1475-1482.
Gullberg, M., & Holmqvist, K. (2006). What speakers do and what addressees look at: Visual attention to gestures in human interaction live and on video. *Pragmatics and Cognition*, *14*(1), 53–82.
Haensel, J. X., Smith, T. J., & Senju, A. (2021). Cultural differences in mutual gaze during face-to-face interactions: A dual head-mounted eye-tracking study. *Visual Cognition*, 1-16.
Hanisch, C., Radach, R., Holtkamp, K., Herpertz-Dahlmann, B., & Konrad, K. (2006). Oculomotor inhibition in children with and without attention-deficit hyperactivity disorder (ADHD). *Journal of neural transmission*, *113*(5), 671-684.
Hanna, J. E., & Brennan, S. E. (2007). *Speakers' eye gaze disambiguates referring expressions early during face-to-face conversation. Journal of Memory and Language, 57(4),* 596–615.
Hanna, J. E., & Tanenhaus, M. K. (2004). Pragmatic effects on reference resolution in a collaborative task: Evidence from eye movements. *Cognitive science*, *28*(1), 105-115.
Hautala, J., Loberg, O., Hietanen, J. K., Nummenmaa, L., & Astikainen, P. (2016). Effects of conversation content on viewing dyadic conversations. *Journal of Eye Movement Research*, *9*.
Hayward, D. A., Voorhies, W., Morris, J. L., Capozzi, F., & Ristic, J. (2017). Staring reality in the face: A comparison of social attention across laboratory and real world measures suggests little common ground. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *71*(3), 212.
Hessels, R. S. (2020). How does gaze to faces support face-to-face interaction? A review and perspective. *Psychonomic Bulletin & Review*, *27*(5), 856-881.
Hessels, R. S., Holleman, G. A., Kingstone, A., Hooge, I. T., & Kemner, C. (2019). Gaze allocation in face-to-face communication is affected primarily by task structure and social context, not stimulus-driven factors. *Cognition*, *184*, 28-43.
Hessels, R. S., Niehorster, D. C., Holleman, G. A., Benjamins, J. S., & Hooge, I. T. C. (2020). Wearable Technology for "Real-World Research": Realistic or Not? In *Perception*, *49 (6)*, 611–615.
Hietanen, J. K. (1999). Does your gaze direction and head orientation shift my visual attention? *NeuroReport*, *10*(16), 3443–3447.
Hirvenkari, L., Ruusuvuori, J., Saarinen, V. M., Kivioja, M., Peräkylä, A., & Hari, R. (2013). Influence of Turn-Taking in a Two-Person Conversation on the Gaze of a Viewer. *PLoS ONE*, *8*(8).
Ho, S., Foulsham, T., & Kingstone, A. (2015). Speaking and listening with the eyes: Gaze signaling during dyadic interactions. *PLoS ONE*, *10*(8), 1–18.
Hoc, J.M. (2000). Toward Ecological Validity of Research on Cognition. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *44*(6), 549–552.
Hockley, W. E., Hemsworth, D. H., & Consoli, A. (1999). Shades of the mirror effect: Recognition of faces with and without sunglasses. *Memory & Cognition*, *27*(1), 128-138.
Hoehl, S., Michel, C., Reid, V. M., Parise, E., & Striano, T. (2014). Eye contact during live social interaction modulates infants' oscillatory brain activity. *Social Neuroscience*, *9*(3), 300-308.
Holleman, G. A., Hessels, R. S., Kemner, C., & Hooge, I. T. (2020). Implying social interaction and its influence on gaze behavior to the eyes. *PloS one*, *15*(2), e0229203.
Holleman, G. A., Hooge, I. T., Kemner, C., & Hessels, R. S. (2020). The 'real-world approach'and its problems: A critique of the term ecological validity. *Frontiers in Psychology*, *11*, 721.

Holler, J., & Kendrick, K. H. (2015). Unaddressed participants' gaze in multi-person interaction: optimizing recipiency. *Frontiers in psychology*, *6*(98), 1-14.
Huang, J. H., & Chan, Y. S. (2020). Saccade eye movement in children with attention deficit hyperactivity disorder. *Nordic journal of psychiatry*, *74*(1), 16-22.
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., & Baldi, P. (2009). Bayesian surprise attracts human attention. *Vision research*, *49*(10), 1295-1306.
Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nature Reviews Neuroscience*, *2*(3), 194–203.
Jarick, M., & Kingstone, A. (2015). The duality of gaze: eyes extract and signal social information during sustained cooperative and competitive dyadic gaze. *Frontiers in psychology*, *6*, 1423.
Jiang, J., von Kriegstein, K., & Jiang, J. (2020). Brain mechanisms of eye contact during verbal communication predict autistic traits in neurotypical individuals. *Scientific reports*, *10*(1), 1-11.
Kajopoulos, J., Cheng, G., Kise, K., Müller, H. J., & Wykowska, A. (2021). Focusing on the face or getting distracted by social signals? The effect of distracting gestures on attentional focus in natural interaction. *Psychological Research*, *85*(2), 491-502.
Kano, F., & Call, J. (2014). Cross-species variation in gaze following and conspecific preference among great apes, human infants and adults. *Animal Behaviour*, *91*, 137-150.
Kanwisher, N., & Wojciulik, E. (2000). Visual attention: Insights from brain imaging. *Nature Reviews Neuroscience*, *1*(2).
Karatekin, C., & Asarnow, R. F. (1998). Components of visual search in childhood-onset schizophrenia and attention-deficit/hyperactivity disorder. *Journal of abnormal child psychology*, *26*(5), 367-380.
Kaspar, K. (2013). What guides visual overt attention under natural conditions? Past and future research. *International Scholarly Research Notices*, *2013*.
Keil, M. S. (2009). "I look in your eyes, honey": Internal face features induce spatial frequency preference for human face processing. *PLoS Computational Biology*, *5*(3), 1000329.
Kelly, D. J., Quinn, P. C., Slater, A. M., Lee, K., Gibson, A., Smith, M., Ge, L., & Pascalis, O. (2005). Three-month-olds, but not newborns, prefer own-race faces. *Developmental Science*, *8*(6).
Kemner, C., Van der Geest, J. N., Verbaten, M. N., & van Engeland, H. (2007). Effects of object complexity and type on the gaze behavior of children with pervasive developmental disorder. *Brain and cognition*, *65*(1), 107-111.
Kendon, A. (1967). Some functions of gaze-direction in social interaction. *Acta Psychologica, 26*, 22–63.
Keppel, G., & Wickens, T. D. (2004). *Design and Analysis: A Researcher's Handbook* (4th ed.). Pearson Prentice Hall.
Kessler, R. C., Adler, L., Ames, M., Demler, O., Faraone, S., Hiripi, E. V. A., ... & Walters, E. E. (2005). The World Health Organization Adult ADHD Self-Report Scale (ASRS): a short screening scale for use in the general population. *Psychological medicine*, *35*(2), 245-256.
Kim, O. H., & Kaiser, A. P. (2000). Language characteristics of children with ADHD. *Communication disorders quarterly*, *21*(3), 154-165.
Kingstone, A. (2009). Taking a real look at social attention. *Current opinion in neurobiology*, *19*(1), 52-56.

Kingstone, A., Kachkovski, G., Vasilyev, D., Kuk, M. & Welsh, T.N. (2019). Mental attribution is not sufficient or necessary to trigger attentional orienting to gaze. *Cognition, 189,* 35-40.
Kingstone, A., Smilek, D., & Eastwood, J. D. (2008). Cognitive ethology: A new approach for studying human cognition. *British Journal of Psychology*, *99*(3), 317-340.
Klein, C., Fischer, B., & Hartnegg, K. (2002). Effects of methylphenidate on saccadic responses in patients with ADHD. *Experimental brain research*, *145*(1), 121-125.
Kleinke, C. L. (1986). Gaze and eye contact: a research review. *Psychological bulletin*, *100*(1), 78.
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., & Kohshima, S. (1997). Unique morphology of the human eye. *Nature*, *387*(6635), 767-768.
Kobayashi, H., & Kohshima, S. (2001). Unique morphology of the human eye and its adaptive meaning: comparative studies on external morphology of the primate eye. *Journal of human evolution*, *40*(5), 419-435.
Koch, C., & Ullman, S. (1985). Shifts in selective visual attention: Towards the underlying neural circuitry. *Human Neurobiology*, *4*(4), 219–227.
Kok, E. M., Aizenman, A. M., Võ, M. L. H., & Wolfe, J. M. (2017). Even if I showed you where you looked, remembering where you just looked is hard. *Journal of Vision, 17(*12), 2-2.
Kovic, V., Plunkett, K,. & Westermann, G. (2009). Eye-tracking study of inanimate objects, *Psihologija, 42*, 417-436.
Kuhn, G., & Benson, V. (2007). The influence of eye-gaze and arrow pointing distractor cues on voluntary eye movements. *Perception & psychophysics*, *69*(6), 966-971.
Kuhn, G., Kourkoulou, A., & Leekam, S. R. (2010). How magic changes our expectations about autism. *Psychological Science*, *21*(10), 1487-1493
Kuhn, G., Tatler, B. W., & Cole, G. G. (2009). You look where I look! Effect of gaze cues on overt and covert attention in misdirection. *Visual Cognition*, *17*(6-7), 925-944.
Kumle, L., Vo, M. L., & Draschkow, D. (2020). Estimating power in (generalized) linear mixed models: an open introduction and tutorial in R. https://doi.org/10.31234/osf.io/vxfbh
Laidlaw, K. E. W., Foulsham, T., Kuhn, G., & Kingstone, A. (2011). Potential social interactions are important to social attention. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(14), 5548–5553.
Lamme, V. A. F. (2003). Why visual attention and awareness are different. *Trends in Cognitive Sciences*, *7*(1), 12–18.
Land, M. F. (2014). *The Eye: A very short introduction*. Oxford University Press.
Langten, S. R. H., Watt, R. J., & Bruce, V. (2000). Do the eyes have it? Cues to the direction of social attention. *Trends in Cognitive Sciences*, *4*(2), 50–59.
Langton, S. R. H., & Bruce, V. (1999). Reflexive visual orienting in response to the social attention of others. *Visual Cognition*, *6*(5), 541–567.
Latif, N., Alsius, A., & Munhall, K. G. (2018). Knowing when to respond: the role of visual information in conversational turn exchanges. *Attention, Perception, and Psychophysics*, *80*(1), 27–41.
Lev, A., Braw, Y., Elbaum, T., Wagner, M., & Rassovsky, Y. (2020). Eye Tracking During a Continuous Performance Test: Utility for Assessing ADHD Patients. *Journal of Attention Disorders*, 1087054720972786
Lorch, E. P., Milich, R., Astrin, C. C., & Berthiaume, K. S. (2006). Cognitive engagement

and story comprehension in typically developing children and children with ADHD from preschool through elementary school. *Developmental Psychology*, *42*(6), 1206.
Macdonald, R. G., & Tatler, B. W. (2013). Do as eye say: Gaze cueing and language in a real-world social interaction. *Journal of Vision*, *13*(4).
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., & Neuberg, S. L. (2003). Sexually Selective Cognition: Beauty Captures the Mind of the Beholder. *Journal of Personality and Social Psychology, 85*(6), 1107–1120.
Mansour, H., & Kuhn, G. (2019). Studying "natural" eye movements in an "unnatural" social environment: The influence of social activity, framing, and sub-clinical traits on gaze aversion. *Quarterly Journal of Experimental Psychology*, *72*(8), 1913-1925.
Maran, T., Furtner, M., Liegl, S., Ravet-Brown, T., Haraped, L., & Sachse, P. (2020). Visual attention in real-world conversation: Gaze patterns are modulated by communication and group size. *Applied Psychology*.
McCarthy, A., Lee, K., Itakura, S., & Muir, D. W. (2008). Gaze Display When Thinking Depends on Culture and Context. *Journal of Cross-Cultural Psychology*, *39*(6), 716–729.
McInnes, A., Humphries, T., Hogg-Johnson, S., & Tannock, R. (2003). Listening comprehension and working memory are impaired in attention-deficit hyperactivity disorder irrespective of language impairment. *Journal of abnormal child psychology*, *31*(4), 427-443
Metzing, C., & Brennan, S. E. (2003). When conceptual pacts are broken: Partner-specific effects on the comprehension of referring expressions. *Journal of Memory and Language*, *49*(2), 201-213.
Mohammadhasani, N., Caprì, T., Nucita, A., Iannizzotto, G., & Fabio, R. A. (2020). Atypical visual scan path affects remembering in ADHD. *Journal of the International Neuropsychological Society*, *26*(6), 557-566.
Morgan, E. J., Foulsham, T., & Freeth, M. (2020). Sensitivity to social agency in autistic adults. *Journal of Autism and Developmental Disorders*, 1-11.
Mostofsky, S. H., Lasker, A. G., Cutting, L. E., Denckla, M. B., & Zee, D. S. (2001). Oculomotor abnormalities in attention deficit hyperactivity disorder: a preliminary study. *Neurology*, *57*(3), 423-430.
Müller, P., Sood, E., & Bulling, A. (2020). Anticipating averted gaze in dyadic interactions. In *ACM Symposium on Eye Tracking Research and Applications*. 1-10.
Mundy, P. (1998). Individual differences in joint attention skill development in the second year. *Infant Behavior and Development*, *21*(3), 469–482.
Mundy, P., & Newell, L. (2007). Attention, joint attention, and social cognition. *Current Directions in Psychological Science*, *16*(5), 269–274.
Munoz, D. P., Armstrong, I. T., Hampton, K. A., & Moore, K. D. (2003). Altered control of visual fixation and saccadic eye movements in attention-deficit hyperactivity disorder. *Journal of neurophysiology*, *90*(1), 503-514.
Myllyneva, A., Ranta, K., & Hietanen, J. K. (2015). Psychophysiological responses to eye contact in adolescents with social anxiety disorder. *Biological Psychology*, *109*, 151-158.
Naselaris, T., Stansbury, D. E., & Gallant, J. L. (2012). Cortical representation of animate and inanimate objects in complex natural scenes. *Journal of Physiology, Paris, 106* (5-6), 239–249.
Nasiopoulos, E., Risko, E. F., Foulsham, T., & Kingstone, A. (2015). Wearable computing:

Will it make people prosocial?. *British Journal of Psychology*, *106*(2), 209-216.
Neider, M. B., Chen, X., Dickinson, C. A., Brennan, S. E., & Zelinsky, G. J. (2010). *Coordinating spatial referencing using shared gaze. Psychonomic Bulletin & Review, 17(5), 718–724.*
New, J., Cosmides, L., & Tooby, J. (2007). Category-specific attention for animals reflects ancestral priorities, not expertise. *Proceedings of the National Academy of Sciences, 104* (42), 16598-16603.
Nguyen, B. L., Chahir, Y., Molina, M., Tijus, C., & Jouen, F. (2010). Eye gaze tracking with free head movements using a single camera. *ACM International Conference Proceeding Series*, 108–113.
Niehorster, D. C., Cornelissen, T. H. W., Holmqvist, K., Hooge, I. T. C., & Hessels, R. S. (2018). What to expect from your remote eye-tracker when participants are unrestrained. *Behavior Research Methods*, *50*(1), 213–227.
Niehorster, D. C., Hessels, R. S., & Benjamins, J. S. (2020). GlassesViewer: Open-source software for viewing and analyzing data from the Tobii Pro Glasses 2 eye tracker. *Behavior Research Methods*, *52*(3), 1244–1253.
Niehorster, D. C., Santini, T., Hessels, R. S., Hooge, I. T., Kasneci, E., & Nyström, M. (2020). The impact of slippage on the data quality of head-worn eye trackers. *Behavior Research Methods*, *52*(3), 1140-1160.
Nielsen, M. K., Slade, L., Levy, J. P., & Holmes, A. (2015). Inclined to see it your way: Do altercentric intrusion effects in visual perspective taking reflect an intrinsically social process? *Quarterly Journal of Experimental Psychology*, *68*(10), 1931–1951.
Norbury, C. F., Brock, J., Cragg, L., Einav, S., Griffiths, H., & Nation, K. (2009). Eye-movement patterns are associated with communicative competence in autistic spectrum disorders. *Journal of Child Psychology and Psychiatry*, *50*(7), 834–842.
Nuku, P., & Bekkering, H. (2008). Joint attention: Inferring what others perceive (and don't perceive). *Consciousness and Cognition*, *17*(1), 339–349.
Ohlsen, G., Van Zoest, W., & Van Vugt, M. (2013). Gender and facial dominance in gaze cuing: Emotional context matters in the eyes that we follow. *PloS one*, *8*(4).
Pascalis, O., De Haan, M., Nelson, C. A., & De Schonen, S. (1998). Long-term recognition memory for faces assessed by visual paired comparison in 3-and 6-month-old infants. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(1), 249.
Penn, C. (2000). Paying attention to conversation. *Brain and Language*, *71*(1), 185–189.
Perea-García, J. O., Kret, M. E., Monteiro, A., & Hobaiter, C. (2019). Scleral pigmentation leads to conspicuous, not cryptic, eye morphology in chimpanzees. *Proceedings of the National Academy of Sciences*, *116*(39), 19248-19250.
Perrett, D. I., Hietanen, J. K., Oram, M. W., & Benson, P. J. (1992). Organization and functions of cells responsive to faces in the temporal cortex. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *335*(1273), 23–30.
Peterson, M. F., & Eckstein, M. P. (2014). Learning optimal eye movements to unusual faces. *Vision Research*, *99*, 57–68.
Posner, M. I. (1980). Orienting of attention. *The Quarterly Journal of Experimental Psychology*, *32*(1), 3–25.
Pratt, J., Radulescu, P. V., Guo, R. M., & Abrams, R.A. (2010). It's Alive!: Animate Motion Captures Visual Attention. *Psychological Science*, *21* (11), 1724-1730.
Purves, D., Augustine, G, J., Fitzpatrick, D., Katz, L, C., LaMantia, A., McNamara, J, O., & Williams, S. M. (2001). *Neuroscience* (2nd ed.). Sinauer Associates.
Quigley, C., Harding, S., Cooke, M., & König, P. (2008). Audio-visual integration during overt visual attention. *Journal of Eye Movement Research*, *1*(2), 1–17.

Riby, D. M., & Hancock, P. J. (2009). Do faces capture the attention of individuals with Williams syndrome or autism? Evidence from tracking eye movements. *Journal of autism and developmental disorders*, *39*(3), 421-431.

Ricciardelli, P., Bricolo, E., Aglioti, S. M., & Chelazzi, L. (2002). My eyes want to look where your eyes are looking: Exploring the tendency to imitate another individual's gaze. *NeuroReport*, *13*(17), 2259–2264.

Richardson, D. C., Altmann, G. T. M., Spivey, M. J., & Hoover, M. A. (2009). Much ado about eye movements to nothing: a response to Ferreira et al.: Taking a new look at looking at nothing. *Trends in Cognitive Sciences, 13* (6), 235–236.

Richardson, D. C., & Spivey, M. J. (2000). Representation, space and Hollywood Squares: Looking at things that aren't there anymore. *Cognition*, *76*(3), 269–295. https://doi.org/10.1016/S0010-0277(00)00084-6

Riggio, R. E., Widaman, K. F., Tucker, J. S., & Salinas, C. (1991). Beauty is more than skin deep: Components of attractiveness. *Basic and applied social psychology*, *12*(4), 423-439.

Risko, E. F., Anderson, N. C., Lanthier, S., & Kingstone, A. (2012). Curious eyes: Individual differences in personality predict eye movement behavior in scene-viewing. *Cognition*, *122*(1), 86-90.

Risko, E. F., Laidlaw, K. E. W., Freeth, M., Foulsham, T., & Kingstone, A. (2012). Social attention with real versus reel stimuli: Toward an empirical approach to concerns about ecological validity. *Frontiers in Human Neuroscience*, *6,* 1–11.

Risko, E, F., Richardson D, C,. & Kingstone, A. (2016). Breaking the Fourth Wall of Cognitive Science: Real-World Social Attention and the Dual Function of Gaze. *Current Directions in Psychological Science*, *25* (1), 70–74.

Ristic, J., Mottron, L., Friesen, C. K., Iarocci, G., Burack, J. A., & Kingstone, A. (2005). Eyes are special but not for everyone: The case of autism. *Cognitive Brain Research*, *24*(3), 715-718.

Roberson, D., Kikutani, M., Döge, P., Whitaker, L., & Majid, A. (2012). Shades of emotion: What the addition of sunglasses or masks to faces reveals about the development of facial expression processing. *Cognition*, *125*(2), 195-206.

Rogers, K. (2011). *The eye, The physiology of human perception*. Britannica Educational Publishing.

Rogers, S. L., Speelman, C. P., Guidetti, O., & Longmuir, M. (2018). Using dual eye tracking to uncover personal gaze patterns during social interaction. *Scientific reports*, *8*(1), 1-9.

Rommelse, N. N. J., Van der Stigchel, S., Witlox, J., Geldof, C., Deijen, J. B., Theeuwes, J., & Sergeant, J. A. (2008). Deficits in visuo-spatial working memory, inhibition and oculomotor control in boys with ADHD and their non-affected brothers. *Journal of neural transmission*, *115*(2), 249-260.

Ross, N. M., & Kowler, E. (2013). Eye movements while viewing narrated, captioned, and silent videos. *Journal of Vision*, *13*(4), 1–1.

Sagiv, N., & Bentin, S. (2001). Structural encoding of human and schematic faces: Holistic and part-based processes. *Journal of Cognitive Neuroscience*, *13*(7), 937–951.

Samson, D., Apperly, I. A., Braithwaite, J. J., Andrews, B. J., & Bodley Scott, S. E. (2010). Seeing it their way: evidence for rapid and involuntary computation of what other people, *Journal of Experimental Psychology: Human Perception and Performance*, *36* (5), 1255.

Sanchez, R. P., Pugzles Lorch, E., Milich, R., & Welsh, R. (1999). Comprehension of televised stories by preschool children with ADHD. *Journal of Clinical Child Psychology*, *28*(3), 376-385.

Santiesteban, I., Catmur, C., Hopkins, S. C., & Bird, G. (2014). Avatars and arrows: Implicit

mentalizing or domain-general processing? *Journal of Experimental Psychology: Human Perception and Performance*, *40*(3), 929–937.
Sayal, K., Prasad, V., Daley, D., Ford, T., & Coghill, D. (2018). ADHD in children and young people: prevalence, care pathways, and service provision. *The Lancet Psychiatry*, *5*(2), 175-186.
Schilbach, L. (2010). A second-person approach to other minds. *Nature Reviews Neuroscience*, *11*(6), 449-449.
Schmuckler, M. A. (2001). What is ecological validity? A dimensional analysis. *Infancy*, *2*(4), 419-436.
Scott, H., Batten, J. P., & Kuhn, G. (2019). Why are you looking at me? It's because I'm talking, but mostly because I'm staring or not doing much. *Attention, Perception, and Psychophysics*, *81*(1), 109–118.
Senju, A., & Johnson, M. H. (2009). Atypical eye contact in autism: models, mechanisms and development. Neuroscience and Biobehavioral Reviews, 33(8), 1204-1214. doi:S0149-7634(09)00082-7 [pii] 10.1016/j.neubiorev.2009.06.001.
Serrano, V. J., Owens, J. S., & Hallowell, B. (2018). Where children with ADHD direct visual attention during emotion knowledge tasks: Relationships to accuracy, response time, and ADHD symptoms. Journal of attention disorders, 22(8), 752-763.
Shepherd, S. V. (2010). Following gaze: gaze-following behavior as a window into social cognition. *Frontiers in integrative neuroscience*, *4*, 5.
Shen, J., & Itti, L. (2012). Top-down influences on visual attention during listening are modulated by observer sex. *Vision Research*, *65*, 62–76.
Sifre, R., Olson, L., Gillespie, S., Klin, A., Jones, W., & Shultz, S. (2018). A Longitudinal Investigation of Preferential Attention to Biological Motion in 2- to 24-Month-Old Infants. *Scientific Reports, 8,* (2527).
Singer, T., & Tusche, A. (2014). Understanding others: Brain mechanisms of theory of mind and empathy. In *Neuroeconomics* (pp. 513-532). Academic Press.
Skripkauskaite, S., Mihai, I., & Koldewyn, K. (2021). Brief report: Attentional Bias towards Social Interactions during Viewing of Naturalistic Scenes. *bioRxiv*.
Spivey, M., Tyler, M., Richardson, D., Young, E. (2000). Eye movements during comprehension of spoken scene descriptions. In: Gleitman, L.R., Joshi, A.K. (Eds.), Proceedings of the Twenty-Second Annual Conference of the Cognitive Science Society, 487–492.
Stiefelhagen, R., & Zhu, J. (2002). Head orientation and gaze direction in meetings. *Conference on Human Factors in Computing Systems - Proceedings*, 858–859.
Suler, J. (2004). The online disinhibition effect. *Cyberpsychology & behavior*, *7*(3), 321-326.
Sumby, W. H., & Pollack, I. (1954). Visual Contribution to Speech Intelligibility in Noise. *Journal of the Acoustical Society of America*, *26*(2), 212–215.
Theeuwes, J., & Van der Stigchel, S. (2006). Faces capture attention: Evidence from inhibition of return. *Visual Cognition*, *13*(6), 657–665.
Teufel, C., Alexis, D. M., Todd, H., Lawrance-Owen, A. J., Clayton, N. S., & Davis, G. (2009). Social cognition modulates the sensory coding of observed gaze direction. *Current Biology, 19* (15), 1274–1277.
Tice, M., & Henetz, T. (2011). The eye gaze of 3rd party observers reflects turn-end boundary projection. *Proceedings of the 15th Workshop on the Semantics and Pragmatics of Dialogue,* 204-205.
Tomasello, M. (1995). Joint attention as social cognition. *Joint attention: Its origins and role in development*, *103130*, 103-130.
Tomasello, M., Hare, B., Lehmann, H., & Call, J. (2007). Reliance on head versus eyes in the gaze following of great apes and human infants: the cooperative eye hypothesis. *Journal*

of Human Evolution, *52*(3), 314–320.
Tsank, Y., & Eckstein, M. (2015). Optimal point of fixation to faces for vision with a simulated central scotoma. *Journal of Vision*, *15*(12), 933.
Vatikiotis-Bateson, E., Eigsti, I. M., Yano, S., & Munhall, K. G. (1998). Eye movement of perceivers during audiovisual speech perception. *Perception and Psychophysics*, *60*(6), 926–940.
Vertegaal, R., & Ding, Y. (2002). Explaining effects of eye gaze on mediated group conversations: amount or synchronization?. In *Proceedings of the 2002 ACM conference on Computer supported cooperative work*, 41-48.
Vertegaal, R., Slagter, R., Van der Veer, G., & Nijholt, A. (2001, March). Eye gaze patterns in conversations: there is more to conversational agents than meets the eyes. In Proceedings of the SIGCHI conference on Human factors in computing systems. 301-308.
Vo, M. L. H., Smith, T. J., Mital, P. K., & Henderson, J. M. (2012). Do the eyes really have it? Dynamic allocation of attention when viewing moving faces. *Journal of Vision*, *12*(13), 3–3.
von dem Hagen, E. A., & Bright, N. (2017). High autistic trait individuals do not modulate gaze behaviour in response to social presence but look away more when actively engaged in an interaction. *Autism Research*, *10*(2), 359-368.
Yarbus, A. L. (1967). *Eye movements and vision*. Plenum Publishing Corporation.
Zhang, L., Tjondronegoro, D., & Chandran, V. (2014). Random Gabor based templates for facial expression recognition in images with facial occlusion. *Neurocomputing*, *145*, 451-464.
Zhu, Z., & Ji, Q. (2007). Novel eye gaze tracking techniques under natural head movement. *IEEE Transactions on Biomedical Engineering*, *54*(12), 2246–2260.
Zion-Golumbic, E., Cogan, G., Schroeder, C., & Poeppel, D. (2013). Visual Input Enhances Selective Speech Envelope Tracking in Auditory Cortex at a "Cocktail Party". *The Journal of neuroscience: the official journal of the Society for Neuroscience. 33.* 1417-26.
Zuckerman, M., Miserandino, M., & Bernieri, F. (1983). Civil inattention exists—in elevators. *Personality and Social Psychology Bulletin*, *9*(4), 578-586.

Appendix

1. Appendix - Experiment 1, 2, 3

Experiment 1 additional analysis

The number of 'Yes' responses were compared to the total number of responses to give overall percentage of 'hit's for each condition.

Overall, regardless of condition, participants coded the fixation was a hit 44.2% of the time. The overall percentage hit for faces were 50.8% and objects 37.8%.

The mean percentages of 'hit's on objects and faces, split by Distance (1-5) and by Cursor Shape and Size can be seen in Table A1 and A2, respectively.

Distance	Small circle	Large circle	Small Cross	Large Cross	Total
1	93.51	93.90	91.56	94.87	93.70
2	73.20	78.33	60.65	67.05	75.76
3	50.17	39.94	43.53	48.18	45.06
4	30.62	20.90	31.69	21.64	25.76
5	23.12	15.25	19.35	17.98	19.19
Total	54.12	49.66	49.36	49.95	50.77

Table A1. Table to show the average 'hit' percentage for images of Faces, split by Distance and Cursor Type.

Distance	Small circle	Large circle	Small Cross	Large Cross	Total
1	87.90	77.93	85.58	74.07	81.37
2	54.59	59.18	50.87	57.59	55.56
3	29.09	24.74	27.40	27.15	27.10
4	17.75	12.19	16.80	10.63	14.34
5	10.33	4.99	12.74	9.93	9.50
Total	39.93	35.81	38.68	35.88	37.57

Table A2. Table to show the average 'hit' percentage for images of Objects, split by Distance and Cursor Type.

Experiment 2 additional analysis

The number of 'Yes' responses were compared to the total number of responses to give overall percentage of 'hits' for each condition.

Overall, regardless of condition, participants coded the cursor as a hit 25.4% of the time. The overall percentage hit rate for images of faces was 26.8% and 23.9% for images of objects. The mean percentages of hits on objects and faces, split by Distance (1-5) be seen in Table A3.

Distance	Faces	Objects	Total
1	95.31	93.27	94.30
2	11.06	8.22	9.57
3	7.39	3.61	5.69
4	5.82	5.53	5.67
5	8.67	4.81	6.68
Total	26.80	23.90	25.35

Table A3. Table to show the average hit percentage for images of Faces and Objects, split by Distance.

Experiment 3 additional analysis

The number of 'Yes' responses were compared to the total number of responses to give overall percentage of 'hits' for each condition.

Overall, regardless of condition, participants coded the cursor as a hit 23.55% of the time. The overall percentage hit rate for images of faces was 27.90% and 19.20% for images of objects. The mean percentages of hits on objects and faces, split by Distance (1-5) be seen in Table A4.

Distance	Faces	Objects	Total
1	96.93	88.02	92.86
2	17.50	8.02	12.92
3	8.05	.95	4.16

4	1.08	1.32	1.21
5	3.29	2.65	2.53
Total	27.90	19.20	23.55

Table A4. Table to show the average hit percentage for images of Faces and Objects, split by Distance.

2. Appendix - Experiment 4

Further method details of Experiment 4

The additional manipulation of eye-contact was removed at a midway point between the two conversations. This was made apparent by the experimenter looking at their watch to signal a change in eye-contact to be used by the researcher at a later stage of analysis. Hence the four procedural conditions were as follows:

Participant uses the mobile eye-tracker walking around campus for roughly 40 minutes. After which they are invited to take part in a conversation which included:

1) Normal eye-contact dyad pair (participant and lead researcher)
2) No eye-contact dyad pair (participant and lead researcher)
3) Normal eye-contact group (participant, lead researcher and confederate)
4) No eye-contact dyad group (participant, lead researcher and confederate)

3. Appendix - Experiment 5

Experiment 5 - additional analysis

Eye tracking experiment

General Viewing Behaviour (eye tracking data)

Table A4 summarizes the average number of fixations and the average fixation duration for each of the four clip conditions.

	Control		Silent		Freeze Frame		Blank	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Number of fixations	102.31	19.03	102.64	22.24	108.57	21.34	90.80	20.32
Fixation duration	391.01	90.74	391.27	97.48	365.50	92.15	434.61	151.24

Table A5. The general measures of oculomotor behaviour averages across participants during each of the four conditions.

Fixations to target faces

Moving regions of interest, in this case the targets faces were then created from the 8 experimental clips. Using Data Viewer (SR research), a dynamic interest area box was drawn around each of the targets faces, which moved throughout the conversation and was logged by slowly playing the clip back with 'mouse record' (an inbuilt function in Data Viewer). Fixations were analysed to determine whether they were inside this area. The three target faces were collapsed to form overall target face location, and this was compared to the x and y coordinates of participant's fixation location.

As a whole, in the 296 trials available for analysis (37 participants with 8 experimental trials each), participants visited all 3 faces of the participants in 96.3% of these trials. In the remaining 3.7% of the fixations, participants visited 2 out of 3 of the targets faces (this is inclusive of all conditions). Therefore, all participants looked at least 2 of the

participant's faces within the 4 conditions, with the vast majority of participants visiting all faces.

Table A5 below shows the percentage of fixations that were on one of the three target faces after clip manipulation. It is noted that the face ROI for the Freeze-Frame and Blank condition is at a different static location to that of the Control and Silent condition. This is due to the fact the face does not move in the Freeze-frame and Control condition; hence the co-ordinates of the face ROI's were at the point of freezing for these conditions. Comparatively, in the Control and Silent condition, the dynamic visual information of the target faces spatial location was still available).

	Control	Silent	Freeze Frame	Blank
Mean percentage of fixations to target faces	87.5	84.8	73.5	23.9
SD	11.1	13.0	19.1	15.6

Table A6. The percentage of fixations on faces, averaged across participants for the 4 conditions, post clip manipulation.

A repeated measures ANOVA established there was a significant difference between conditions, F, (3,108) = 249.36, $p<.001$. Post-hoc analysis with the Bonferroni correction, revealed there were no significant differences between the Control and Silent condition. Hence, the looks to faces did not significantly decrease with the sound muted. All other conditions were significantly different.

Analysis per target member

Further analysis was then carried out to address a potential criticism of the 'central bias'. This analysis was included to ensure participants fixations were not solely in the centre of the screen (as is a common phenomenon in visual attention research) and hence on the middle target. The percentage of fixations spent fixating to the target speaker (whole target ROI) post-clip manipulation is shown in Table A6 and split by condition in Figure A1.

	Target 1	Target 2	Target 3	Elsewhere
Mean percentage of fixations	25.4	41.3	23.3	9.9

Table A7. Table to show the percentage of fixations to target speakers across all clips.

As 'elsewhere' on the screen acquires greater than 25% of the whole visible area, it is apparent that targets were expectedly more attractive to participants. This is not surprising, as we would expect participants to attend to the social stimuli within a scene. Furthermore, the targets are dynamic in comparison to the static background, which again attracts more visual attention.

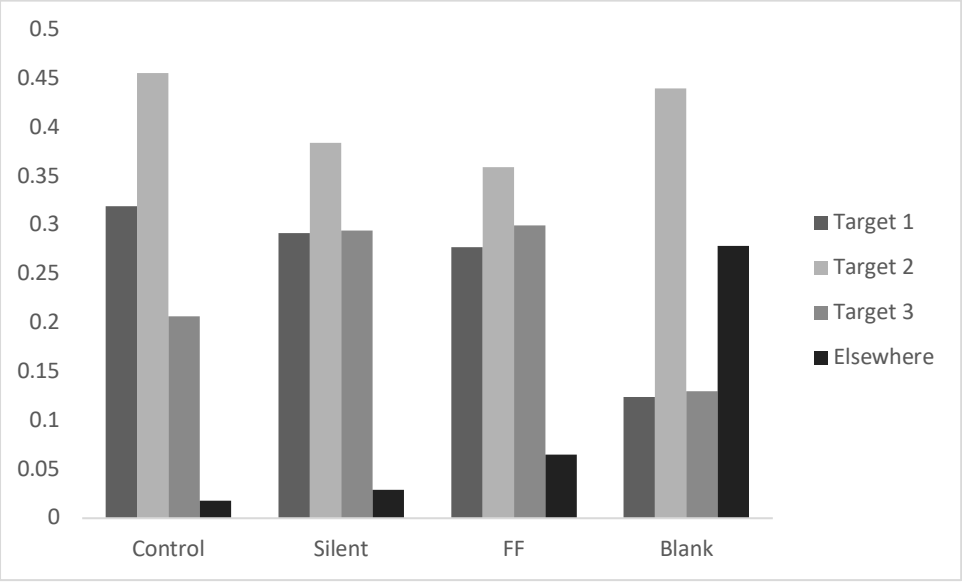

Figure A1. Graph to show percentage of fixations on each target for each condition.

Live behaviour

Visual attention to speaking targets

When looking only at percentage time on speaking targets when the target observers themselves are not speaking and hence take on the role of the listener, the overall looks to speakers average was 53.4%. Hence, overall, 53.4% of the time, when another target was speaking, the target was looking at the speaker.

This is split by target number (4-6) in Table A7. It is important to note that this target split is an average taken from 4 targets (as there were 4 groups (i.e., Target 4 consists of 4 target observers)).

Gaze location	Target 4 (N=4)	Target 5 (N=4)	Target 6 (N=4)	All targets combined
On a speaking target (%)	56.28	50.00	54.03	53.34
On a non-speaking target (%)	34.29	28.15	26.78	29.74
Elsewhere (%)	3.26	4.70	8.00	5.32

Table A8. Table to show the average percentage time spent in each of the gaze locations, split by target number.

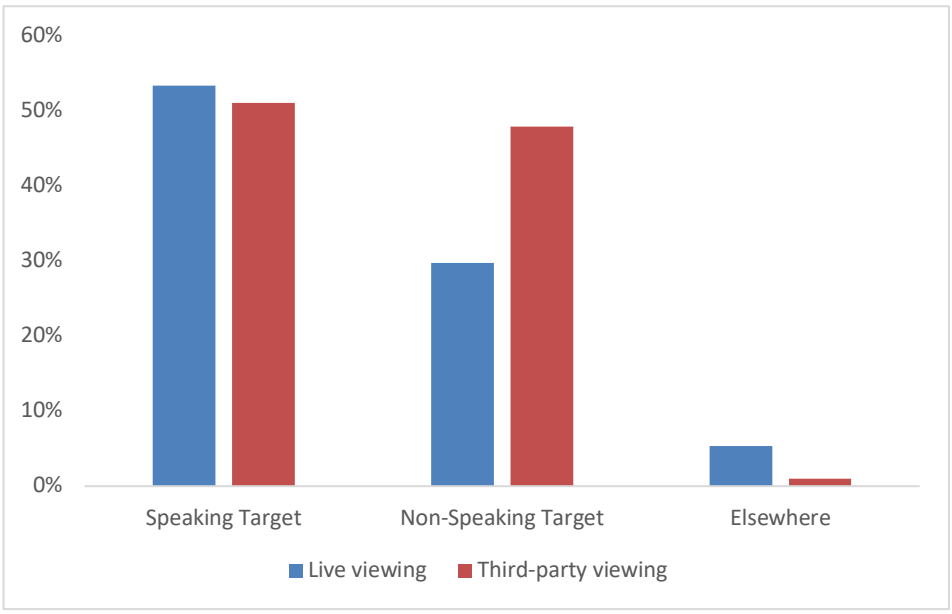

Figure A2. Demonstrates the % time spent on a speaking target, on a target who is not currently speaking, and elsewhere.

Live and third-party agreement

Additional analysis of percentage agreement between the targets and participants can be seen in table A8 below.

		T4	T5	T6
P1	*% agreement*	71%	82%	81%
	κ	.507	.693	.663
	p	*<.001*	*<.001*	*<.001*
P2	*% agreement*	65%	74%	72%
	κ	.450	.586	.542
	p	*<.001*	*<.001*	*<.001*
P3	*% agreement*	60%	74%	70%
	κ	.370	.592	.505
	p	*<.001*	*<.001*	*<.001*
P4	*% agreement*	64%	76%	76%
	κ	.410	.602	.596
	p	*<.001*	*<.001*	*<.001*
P5	*% agreement*	63%	74%	73%
	κ	.409	.572	.548
	p	*<.001*	*<.001*	*<.001*
P6	*% agreement*	63%	78%	74%
	κ	.409	.648	.574
	p	*<.001*	*<.001*	*<.001*
P7	*% agreement*	65%	74%	75%
	κ	.440	.570	.574
	p	*<.001*	*<.001*	*<.001*
P8	*% agreement*	64%	76%	73%
	κ	.440	.623	.566
	p	*<.001*	*<.001*	*<.001*

Table A9. The percentage agreement in which target was being looked at, with Cohens kappa value and significance level.

4. Appendix - Experiment 6

Experiment 6 – additional plots

Here I present two additional plots which highlight the patterns of looking to targets and looking to speaking targets. Both plots display the same data in different graphs.

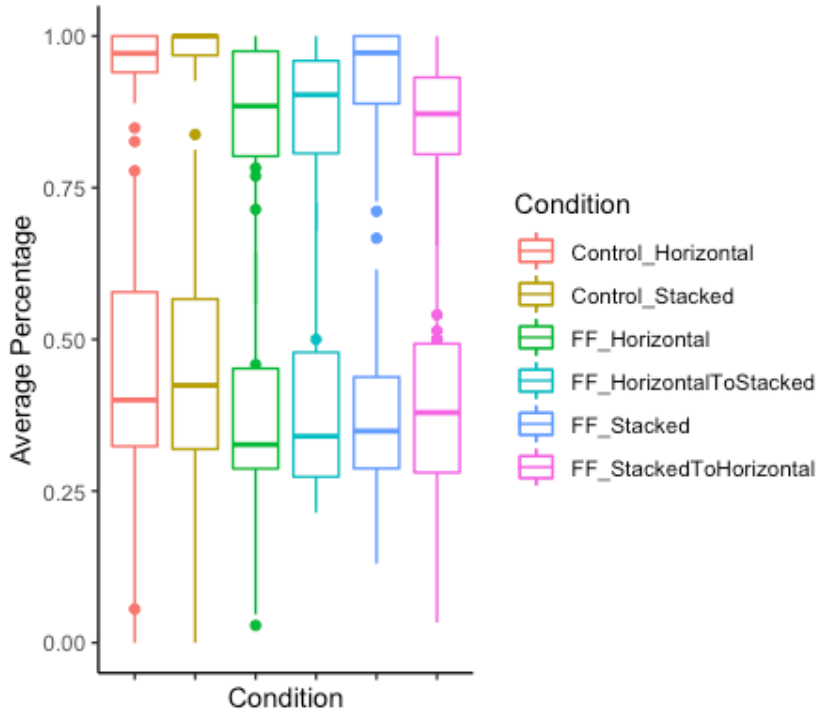

Figure A3. Demonstrates fixations to targets as a whole (top box plots) and fixations to speakers (lower box plots) split by the six conditions. Boxes show the median and quartiles with outliers represented as dots beyond.

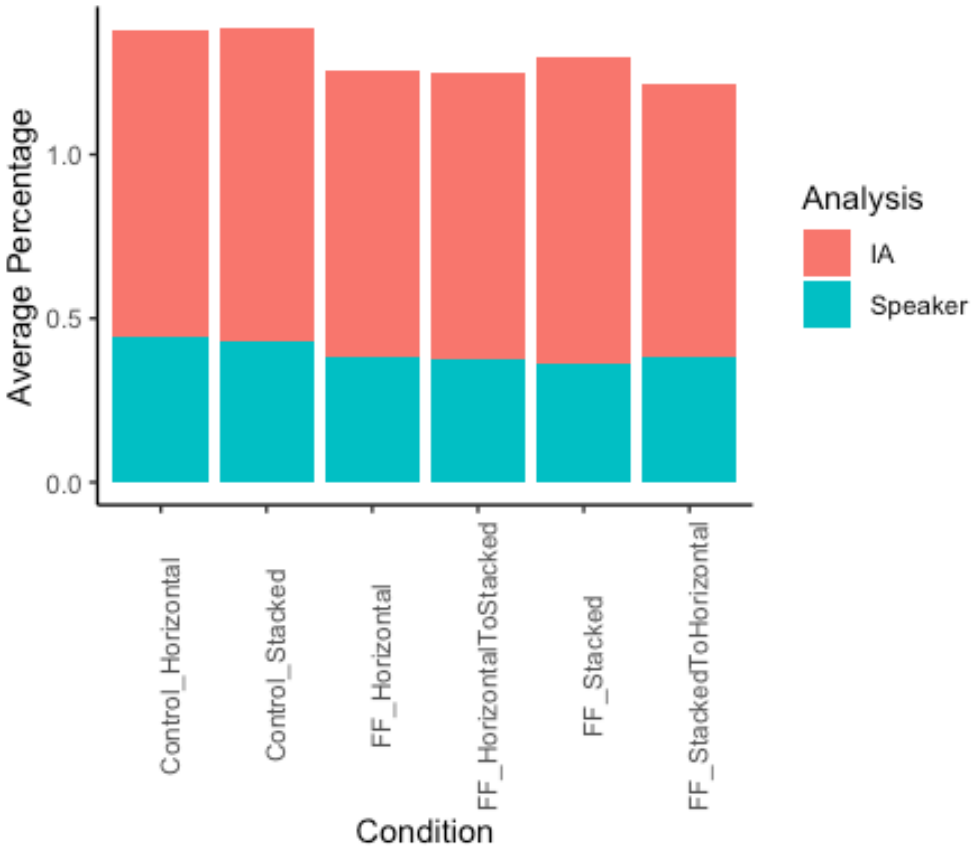

Figure A4. Demonstrates fixations to targets interest areas as a whole (red bars) and fixations to speakers (blue bars) split by the six conditions.

5. Appendix – Research Proposal

Research proposal: What are we attending to when we avert our eyes during live conversation?

The study presented next is a planned study which could not go ahead due to the global pandemic of COVID-19 (see start of thesis for an impact statement). Despite this, I present the relevant methods and hypothesis with the view to complete this research when able in the future. This planned experiment tackles the overall questions of 'where do we attend to when averting our eyes during speaking 'and 'how does eye-contact modulates this'?

Planned MET Proposal

Planned aims

In this study I am interested in why people avert their eyes during conversation, particularly when speaking. We will use a mobile eye-tracker to:

1) Confirm the times during a conversation when participants avert their eyes.
2) Examine where the participant gazes when averting their eyes.
3) Examine whether participants remember the locations they fixated while averting their gaze.
4) Explore what effect occluding the experimenter's eyes has on gaze in conversation.

Overall, I will investigate where participants look when they avert their eyes during speaking and whether they are paying attention to these locations. The study will help us to better understand why we avert gaze during conversation and whether we are attending to objects in our environment or whether this eye movement is just a social signal used in conversation.

Planned methods

Participants will wear a mobile eye-tracker whilst walking around the psychology building. Once they have completed a short walk, we will ask them to enter a room in the lab where there will be objects/posters strategically located. The participants will then orally answer questions regarding how they felt wearing the mobile eye-tracker. For example, 'was the device comfortable to wear?'.

The walk around the building and their question responses will not be used in analysis. The critical period which will be examined will be during the participants conversation with the experimenter in the lab.

An additional manipulation will be occluding the experimenter's eyes during this conversation with sunglasses. The experimenter will wear sunglasses for half of the testing session. After the conversation has finished, the participants will be taken out of the lab room and asked to recall the objects/posters. They will not be aware of this beforehand. The testing session, including set up is estimated to last no longer than 30 minutes. Each participant will take part in 1 session.

Hypotheses and predictions

1) Confirm the times during a conversation when participants avert their eyes.
    a. I hypothesize participants will avert their eyes when they begin speaking in line with previous work by Ho, Foulsham and Kingstone (2015). Here I expect to find a similar pattern.
2) Examine where the participant gazes when averting their eyes.
    a. I hypothesize participants will avert their eyes to areas of the environment which are not occupied by salient or complex objects (i.e., a plain wall) more

so than salient areas (i.e., a busy poster). It is hypothesized this may help to gather thoughts.

3) Examine whether participants remember the locations they fixated while averting their gaze.
    a. I predict participants will not remember the details of the locations they fixated while averting their eyes. I hypothesis this aversion isn't to visually explore or process the environment, but instead is a social signal.
4) Explore what effect occluding the experimenter's eyes has on gaze in conversation.
    a. I expect occluding the experimenter's eyes with sunglasses will result in a less synchronous pattern of behaviour both on an inter-individual and intra-individual level.