Gaze selection in the real world: finding evidence for a preferential selection of eyes.

by

ELINA BIRMINGHAM

B. Sc., University of British Columbia, 2002
M. A., University of Toronto, 2003

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE STUDIES

(Psychology)

THE UNIVERSITY OF BRITISH COLUMBIA

July, 2008
© Elina Birmingham, 2008
Abstract

We have a strong intuition that people's eyes are unique, socially informative stimuli. As such, it is reasonable to propose that humans have developed a fundamental tendency to preferentially attend to eyes in the environment. The empirical evidence to support this intuition is, however, remarkably thin. Over the course of eight chapters, the present thesis considers the area of social attention, and what special role (if any) the selection of eyes has in it. Chapters 2 and 3 demonstrate that when observers are shown complex natural scenes, they look at the eyes more frequently than any other region. This selection preference is enhanced when the social content and activity in the scene is high, and when the task is to report on the attentional states in the scene. Chapters 4 and 5 establish that the bias to select eyes extends to a variety of tasks, suggesting that it may be fundamental to human social attention. In addition, Chapter 5 shows that observers who are told that they will have to remember the scenes look more often at the eyes than observers who are unaware of the forthcoming memory test; moreover this difference between groups persists to scene recognition. Chapter 6 examines whether the preference for eyes can be explained by visual saliency. It cannot. Chapter 7 compares the selection of eyes to another socially communicative cue, the arrow. The results shed light on a recent controversy in the social attention field, and indicate again that there is a fundamental bias to select the eyes. Collectively the data suggest that for typically developing adults, eyes are rich, socially communicative stimuli that are preferentially attended to relative to other stimuli in the environment.
Table of Contents

Abstract ......................................................................................................................................... ii
Table of Contents ......................................................................................................................... iii
List of Tables ................................................................................................................................. v
List of Figures ............................................................................................................................... vi
Acknowledgments ....................................................................................................................... viii
Dedication .................................................................................................................................... ix
Co-authorship statement ............................................................................................................... x

CHAPTER 1
General Introduction .................................................................................................................... 11
  Gaze Perception and Social Attention .................................................................................... 12
  Models of Social Attention ..................................................................................................... 12
  Selection and spatial orienting (shift) components of visual spatial attention ....................... 14
  Spatial orienting (shift) to gazed-at locations ........................................................................ 17
  Selection of gaze cues ............................................................................................................ 19
  Thesis overview ...................................................................................................................... 25

CHAPTER 2
Social Attention and real world scenes: .................................................................................... 33
the roles of action, competition, and social content. ................................................................ 33
  Method ................................................................................................................................... 38
  Results ................................................................................................................................... 39
  Discussion ............................................................................................................................... 42
  References .............................................................................................................................. 52

CHAPTER 3
Gaze selection in complex social scenes .................................................................................... 55
  Method ................................................................................................................................... 59
  Results ................................................................................................................................... 61
  Discussion ............................................................................................................................... 63
  References .............................................................................................................................. 73

CHAPTER 4
Is there a default bias to look at the eyes? .................................................................................. 75
  Method ................................................................................................................................... 80
  Results ................................................................................................................................... 81
  Discussion ............................................................................................................................... 84
  References .............................................................................................................................. 93

CHAPTER 5
Remembering social scenes ....................................................................................................... 95
  Method ................................................................................................................................... 98
  Results .................................................................................................................................. 100
  Discussion ............................................................................................................................. 103
  References ............................................................................................................................ 111

CHAPTER 6
Saliency does not account for fixations to eyes within social scenes ....................................... 113
  Methods ................................................................................................................................. 117
  Results .................................................................................................................................. 118
List of Tables

CHAPTER 6
Table 6.1. Median values for saliency of fixated regions, uniform-random saliency, and biased-random saliency, as a function of experiment. ................................................................. 124
List of Figures

CHAPTER 1
Figure 1.1. The salient regions of a face ................................................................. 27
Figure 1.2. Scanpaths of an individual face (A), a face accompanied by the rest of the body (B), and a social scene (C) (Yarbus, 1967) .............................................................................................. 28

CHAPTER 2
Figure 2.1. Scanpaths of an individual face (A), a face accompanied by the rest of the body (B), and a social scene (C) (Yarbus, 1967) .............................................................................................. 47
Figure 2.2. A. Examples of the four scene types. B. Corresponding regions of interest used in analysis. C. Corresponding plots of fixations for all subjects ................................................................. 48
Figure 2.3. Fixation proportion data plotted as a function of People, Activity, and Region ................. 49
Figure 2.4. Cumulative fixation proportions for 1-person scenes and 3-people scenes ........................... 50
Figure 2.5. Duration proportions as a function of People, Activity, and Region ................................ 51

CHAPTER 3
Figure 3.1. Scanpaths of an individual face (A), a face accompanied by the rest of the body (B), and a social scene (C) (Yarbus, 1967) .............................................................................................. 68
Figure 3.2. A. Examples of the four scene types. B. Corresponding regions of interest used in analysis. .............................................................................................................................................. 69
Figure 3.3. Fixation proportion data plotted as a function of Task and Region .................................. 70
Figure 3.4. Proportion of second fixations falling on each region as a function of task (Look, Describe, Social Attention) .......................................................................................................... 71
Figure 3.5. Fixation proportion data plotted as a function of People, Activity, and Region .................... 72

CHAPTER 4
Figure 4.1. Scanpaths of an individual face (A), a face accompanied by the rest of the body (B), and a social scene (C) (Yarbus, 1967) .............................................................................................. 88
Figure 4.2. A. Examples of the four scene types. B. Corresponding regions of interest used in analysis. .............................................................................................................................................. 89
Figure 4.3. Fixation proportion data plotted as a function of Instruction and Region ............................ 90
Figure 4.4. Proportion of second fixations landing in each region as a function of task (Attention, Emotion, Cognitive) ............................................................................................................. 91
Figure 4.5. Cumulative fixation proportions eyes as a function of viewing interval (1-s bins) and Instruction ........................................................................................................................................ 92

CHAPTER 5
Figure 5.1. A. Examples of the four scene types. B. Corresponding regions of interest used in analysis. .............................................................................................................................................. 107
Figure 5.2. Fixation proportions for the People scenes as a function of Instruction (Told, Not Told), Session (Study, Test), and Region .................................................................................... 108
Figure 5.3. Fixation proportions for the No People scenes as a function of Instruction (Told, Not Told), Session (Study, Test), and Region .................................................................................... 109
Figure 5.4. Proportion of first fixations landing in each region of People scenes as a function of Instruction and Session ............................................................................................................. 110

CHAPTER 6
Figure 6.1. Original saliency model, adapted from Koch & Ullman (1985) ............................................ 125
Figure 6.2. Examples of the scenes used and their corresponding saliency maps, overlaid with the first fixations of Experiment 1 ............................................................................................................. 126
CHAPTER 7
Figure 7.1. The scenes used in the experiment. (A) Scenes with eyes and arrows, (B) Scenes with large arrows, (C) Scenes without people ........................................................................... 150
Figure 7.2. Images overlaid with all participants’ fixations. .......................................................... 149
Figure 7.3. Fixation proportions for the scenes with eyes and arrows............................................. 150
Figure 7.4. Fixation proportions for the scenes with large arrows.................................................... 151
Figure 7.5. Fixation proportions for the scenes with no people........................................................ 152
Figure 7.6. Proportion of first fixations in the scenes with eyes and arrows.................................... 153
Figure 7.7. Proportion of first fixations in scenes with large arrows............................................... 154
Figure 7.8. Proportion of first fixations in the scenes with no people............................................. 155
Acknowledgments

First, I would like to give my warmest thanks to my supervisor, Alan Kingstone, for his continued guidance, inspiration, and support. I feel honoured to have had such an exceptional advisor, and I look forward to many future collaborations. I would also like express my sincere gratitude to Walter Bischof, who has provided invaluable support, guidance, and humour, throughout my graduate training. Of course, I would also like to thank other important people in my life, my parents, Britt and Alan, and my fiancé Michael, for always supporting me in my career goals and encouraging me to pursue my dreams. The research reported in this dissertation was funded by junior and senior graduate fellowships from the Natural Sciences and Engineering Research Council of Canada (NSERC) and the Michael Smith Foundation for Health Research (MSFHR). Additional support came from grants awarded to Alan Kingstone from NSERC and MSFHR.
Dedication

I would like to dedicate this work to my grandmother, Dr. Marion Birmingham. Her pep for life and passion for science has inspired me in my graduate training, and in life. She has taught me to never stop learning.
Co-authorship statement

I am the primary author on all the work presented in this PhD Dissertation, and was responsible for the design of experiments, data collection, data analysis, and manuscript preparation.
CHAPTER 1

General Introduction
Gaze Perception and Social Attention

Imagine the following scenario. You are sitting in a busy cafe and notice that there is a man who is gazing at something behind you. Using his gaze direction you turn around to see what he is looking at. As this simple example illustrates, folk knowledge suggests that we are very interested in the attention of other people. The intuition that we care about the attentional states of others has led to the birth of research in social attention.

While there are several cues to the direction of another person’s attention (e.g., gaze direction, head position, body position, pointing gestures), as the above example suggests, eyes may have a special status as social attentional cues (Baron-Cohen, 1994; Emery, 2000; Langton et al., 2000; Perrett et al., 1992). Morphologically, the human eye is highly salient within the face (Emery, 2000; Kobayashi & Koshima, 1997) and may have evolved specifically to communicate social information to conspecifics (Emery, 2000). For instance, the high iris-to-sclera contrast unique to the human eye promotes fast discrimination of gaze direction at relatively large distances (Kobayashi & Koshima, 1997), providing a reliable, silent indicator of where someone is attending. Humans are not only very accurate at discriminating gaze direction (Cline, 1967; Gibson & Pick, 1963; Lord & Haith, 1974; Perrett & Milders, 1992), but we also appear to have neural structures that are preferentially biased toward processing gaze information. For instance, single cell recordings in monkeys show that the superior temporal sulcus (STS) has cells that are selective for different gaze directions, independent of face orientation (Perrett et al., 1985); and neuroimaging studies (Hoffman & Haxby, 2000; Pelphrey et al., 2004) have similarly shown that the human STS seems to be especially activated by changes in gaze direction (although finding selectivity for specific gaze directions has been inconsistent (George et al., 2001; Kawashima et al., 1999; Wicker et al., 1998)

Models of Social Attention

The intuition that eyes are unique social stimuli is reflected in models of social attention that incorporate eye gaze direction as the primary cue to the social attention of others. The two most prominent models are those put forward by Simon Baron-Cohen and David Perrett.
Baron-Cohen’s model (1994, 1995), posits that an ‘eye direction detector’ (EDD) is a key component of a ‘mind-reading system’ that makes mental state attributions about other people. This system also includes an ‘intentionality detector’ (ID), a ‘shared-attention mechanism’ (SAM) and a ‘theory-of-mind mechanism’ (ToMM). In the mind-reading system, the EDD, which is assumed to be developed by 9 months of age, is a module (Fodor, 1983) that rapidly and obligatorily detects the presence of eyes and computes eye direction.

These two functions of gaze detection and gaze direction computation are extremely important because they are assumed to be the prerequisites to establishing a state of shared attention (often referred to as ‘joint attention’). Shared attention is defined as the understanding that you and another organism are attending to the same thing. This joint attention is proposed to be performed by the SAM, and is thought to be developed between 9 and 18 months of age. The SAM is thought to serve several functions: (a) shift attention in same direction as another person’s gaze (as computed by the EDD); (b) identify that you and the other person are attending to the same thing; (c) link states of ‘seeing’ (from the EDD) to intentions and desires (from the ID); and (d) feed this information into the ToMM (which is assumed to be developed after 18 months of age) so that complex attributions of mental state can be made (e.g., pretend, know, think, believe, etc.). Thus, in this model the eyes are not only important indicators of another person’s attention, they are the key input into a system that is responsible for the development of theory of mind.

David Perrett’s model focuses on how the brain processes the social attention of others. Perrett et al. (1992) propose that in the hierarchy of social communicative cues, eye gaze is the primary cue to the direction of another person’s attention. Perrett notes that while other cues like head and body position can signal the direction of another person’s attention, and so feed into a general ‘direction-of-attention-detector’ (DAD), eye gaze information overrides all these other cues. Thus, if conflicting cues create ambiguity about the direction of social attention (e.g., eye gaze points upward but the head points downward), this is solved by inhibitory signals from eye gaze information to the competing DAD inputs from other cues.
Both of these models place eye gaze as the most important cue to social attention (Baron-Cohen, 1995; Perrett et al., 1992), with the former claiming additionally that processing gaze is supported by an eye-direction detector (EDD) module, the functioning of which is critical to the development of higher functions like theory of mind (Baron-Cohen, 1994).

Implicit in these models, and in the example given at the beginning of the introduction, is that there are at least two distinct stages to social attention. First, we select the eyes as a key social stimulus (e.g., EDD), and second, we shift our attention from the eyes of a person to the location/object that the person is looking at (e.g., SAM). Indeed, current theories of social attention have considered these two stages to be obligatory in nature (e.g., Baron-Cohen, 1994). That is, humans have evolved to preferentially select eyes from their environment, and to shift their attention to where those eyes are looking. Before moving on to a literature review of research on selection of eyes and spatial orienting to gazed-at locations, I will consider the theoretical constructs of selection and spatial orienting of attention.

**Selection and spatial orienting (shift) components of visual spatial attention**

The basic assumption behind theories of visual selective attention is that the visual system is limited in capacity, in that it can only process a relatively small number of items at a time. The implication of this capacity limitation is that we must select some items for processing at the expense of others (hence the term visual selective attention). Several theories of visual selective attention were born from the early work of Broadbent (1958), who proposed a filter theory based on his work with auditory attention. An overlapping component of many of these theories is that visual processing occurs in two distinct stages (Cave & Wolfe, 1990; Duncan, 1980; Duncan & Humphreys, 1989; Treisman & Gelade, 1980; Treisman, 1988; Treisman & Sato, 1990; Wolfe et al., 1989). The first, preattentive, stage (Neisser, 1967) consists of unlimited parallel processing of all stimuli in the visual field. In this preattentive stage, attention is not focused on any individual location in the display, and thus this stage is considered to occur before selection. The second, attentive, stage of processing involves focused attention, and is limited in capacity. In this second stage, attention is focused on individual items in the display, and thus this stage is thought to operate as a result of selection. Thus, the selection of
objects for further processing in the second stage logically occurs as a result of preattentive processing of the visual field. Therefore, items that are passed from the first stage to the second stage of processing are said to be selected (Theeuwes, 1993). However, selection can be thought of more generally as what happens whenever spatial attention is focused on an item/location (Theeuwes, 1993).

Space-based theories claim that visual selective attention is inherently spatial in nature, i.e., locations containing objects are selected. Furthermore, there are two components of attention: selection, which is the process by which some items are chosen for further processing and others are ignored; and spatial orienting, whereby focal attention is shifted from one spatial location to another. These components are intricately linked, for instance, shifting focused spatial attention to a location is thought to be the operation by which information in the visual field is selected for further processing.

How does this attentional focusing occur? One popular metaphor is that attention may be likened to a spotlight “shining” on selected locations, with items within the spotlight’s beam of attention benefiting from enhanced perceptual processing (e.g., Broadbent 1982; Downing & Pinker, 1985; Eriksen & Hoffman, 1973; Posner, 1980; Shulman et al., 1979; Tsal, 1983). This spotlight is limited in capacity, that is, in spatial extent (Yantis, 1988), which is the underlying reason for the capacity limitations in the second, attentive, stage of visual processing. Thus, this spotlight must shift in order to attend to items at different locations in the display (spatial orienting).

Two modes of orienting may be identified: exogenous (or reflexive) orienting and endogenous (or voluntary) orienting. Exogenous orienting is said to occur when attention is captured obligatorily by an external stimulus. A typical instance of this occurs when attention is drawn to a salient stimulus event in the visual periphery, such as when there is an abrupt stimulus onset or luminance change. Endogenous orienting is said to occur in response to internal goals or expectancies of the individual regarding the visual world. For instance, if one knows that a target is likely to appear on the left side of the screen, one can voluntarily orient their attention there in anticipation of the target. When a shift of attention occurs by moving the
eyes (as in most everyday situations) it is called “overt orienting”, and when this shift occurs without the eyes moving, it is called “covert orienting”. Although there are certainly different viewpoints on the link between covert and overt orienting, there is broad agreement that covert and overt orienting are coupled, and without fail all researchers seem to accept the idea that covert shifts of attention precede and are directed at the location of a subsequent overt shift of attention (Deubel & Schneider, 1996; Henderson, 1992; Hoffman & Subramanian, 1995; Kowler et al., 1995; Shepard et al., 1986). Thus, eye movements may be considered to be the outcome of an attentional selection process that occurs before the eye movement itself (Theeuwes, 1993). For the purpose of the literature review in this dissertation, I will consider studies of covert and overt orienting together.

A popular paradigm for studying the movement of the spotlight of attention is the Posner cuing paradigm. (This paradigm will resurface later when I discuss studies of reflexive orienting to gazed-at locations in Chapter 7). In this paradigm, the subject is asked to fixate the centre of a display containing a fixation point flanked on either side by two empty boxes. The task on each trial is to detect the presence of a visual target which can appear inside either box. At the beginning of the trial, a “cue” appears, directing attention to one of the boxes. To study endogenous orienting of attention, the cue is normally an arrow appearing in the centre of the screen and pointing to the left or to the right. Importantly, the arrow predicts the correct location of the target on the majority of trials (predictive central cue). To study exogenous orienting of attention, the attention cue is the sudden brightening of one of the peripheral boxes. Here, the spatial relation between the cue and target is random in nature (spatially nonpredictive peripheral cue) – in other words, the cue does not predict the location of the target and subjects are typically told to ignore the cue altogether. The target then appears at either the cued location or the uncued location. On neutral trials, no location is cued at all, e.g., the arrow is replaced by a plus-sign or the brightening occurs at both locations simultaneously.

The results of a cost-benefit analysis of the response time data show that target detection is speeded on valid trials (when the target location appears at the cued location) relative to neutral trials. In contrast, target detection is slowed on invalid trials (when the target
appears at the uncued location) relative to neutral trials. The explanation for these cuing effects is relatively straightforward. In the case of the predictive central cue, it is inferred that attention is allocated voluntarily to the location pointed to by the arrow. In the case of the nonpredictive central cue, it is inferred that attention is drawn reflexively by the brightening of the box because the brightening does not predict target location and thus there is no incentive for participants to attend to the cued location. In addition, it is thought that the reason target detection is slowed on invalid trials relative to neutral trials is because in order to detect a target at the uncued location, attention must be shifted away from the cued location and towards the uncued location to select the target.

**Spatial orienting (shift) to gazed-at locations**

To date social attention research has focused on the shift stage of social attention by examining the extent to which gaze direction can trigger an attention shift in others. These studies have established that infants (Hood, Willen & Driver, 1998), preschool children (Ristic, Friesen & Kingstone, 2002) and adults (Driver et al., 1999; Friesen & Kingstone, 1998; Langton & Bruce, 1999) will shift their attention automatically to where others are looking. Some of these seminal 'gaze-cueing' studies with adults will be reviewed in the following paragraphs.

In the late 1990’s, a flurry of research emerged that demonstrated obligatory, or reflexive, shifts in attention toward gazed at locations. One of the first studies was conducted by Friesen and Kingstone (1998). Modifying the standard Posner cuing paradigm, the authors replaced the typical central arrow cue with a schematic face. On each trial the eyes could either look left, right, or straight ahead (neutral trials). After a variable interval, a target (F or T) appeared to either the left or the right of the face, and the task was to either detect the target, localize the target (left or right), or identify the target. Importantly, direction of gaze was uninformative, in that it did not predict where the target would occur. In all conditions, response times were faster to targets on cued trials (target appeared at gazed-at location) than on uncued or neutral trials. In addition, the cuing effect emerged early on, within 100 ms post-cue. The conclusion was that attention was shifted reflexively in the direction of gaze. This was a very unique finding in the context of the attention literature, which in the past had found that central
nonpredictive cues did not produce reflexive orienting (Jonides, 1981). Until Friesen and Kingstone’s study with nonpredictive central gaze cues, it had always been assumed that central cues had to be spatially predictive in nature to produce an orienting effect (which was therefore necessarily voluntary in nature). The discovery that central nonpredictive gaze cues could induce reflexive orienting was thought to reflect that eyes are unique, socially and biologically important stimuli (Friesen & Kingstone, 1998).

This basic ‘gaze-cueing’ result has been replicated many times. Langton and Bruce (1999) manipulated the head direction of photographed faces (always with congruent gaze) and found reflexive cuing effects in the direction the head/gaze was pointing. Other studies have demonstrated that the reflexive gaze-cuing effect persists across different facial expressions and across schematic and photographed faces (Hietenan & Lappanen, 2003). In addition, the effect is robust. Driver et al. (1999) showed photographs of real faces and found similar RT benefits for targets appearing at the gazed-at location, even when participants were told that the target was four times as likely to appear at the other location.

Is the attentional orienting in response to gaze cues truly reflexive in the traditional sense of the word? Consistent with previous studies of reflexive attention to non-biological cues (e.g., abrupt peripheral onsets), the gaze cueing effect appears early (100-300 ms) and is relatively short-lived, lasting only about 1 second (Kingstone et al., 2000; Langdon & Smith, 2005). However, it has been suggested that the gaze-cueing effect is a special type of reflexive orienting, different from the reflexive orienting seen for the typical non-biological cues. First, in all of the gaze-cueing studies there has been an absence an inhibitory after-effect called inhibition of return (IOR). With peripheral onset cuing, the initial RT benefit at the cued location relative to the uncued location (facilitation) is replaced after approximately 300 ms with an RT cost at the cued location relative to the uncued location (IOR). This cross-over effect has been considered to be a hallmark of reflexive attention (Tipper & Kingstone, 2005), and the absence of IOR in gaze-cueing tasks suggests that orienting in response to gaze cues may reflect a unique type of reflexive attention (Kingstone et al., 2000; Friesen & Kingstone, 2003b; but see Frischen & Tipper, 2004, for evidence for an late ‘inhibitory’ effect revealed by reorienting
Attention back to screen centre). Second, there is evidence that the reflexive orienting to gaze cues is independent of reflexive inhibition to abrupt onset cues. For example, Friesen and Kingstone (2003b) showed that reflexive orienting to a gazed-at location overlapped in time with inhibition of a location cued by an abrupt onset. Third, there is evidence that the reflexive gaze-cueing effect is special in that it is independent from volitional orienting. In a counterpredictive cueing paradigm, Friesen, Ristic and Kingstone (2004) presented a central schematic face that could look at one of four possible target locations, but participants were informed that the target was 75% likely to appear at the location opposite to gaze direction, with the three non-predicted locations each receiving 8% of targets. The key finding was that Friesen et al. found both a reflexive orienting of attention to the gazed-at location and a volitional orienting effect at the predicted location, relative to the baseline locations. Thus, the authors suggested that the two types of orienting occurred independently and concurrently. Fourth, there is evidence from a study with split-brain patients that the reflexive gaze-cueing effect is lateralized to the hemisphere specialized for face processing (Kingstone et al., 2000). However, split-brain patients show no lateralization of reflexive orienting to nonpredictive central arrows, with the cueing effect occurring in both hemispheres (Ristic, Friesen, & Kingstone, 2002). Similarly, Akiyama et al. (2006) found that a patient with damage to her right superior temporal gyrus (STG) showed no orienting in response to gaze cues but preserved orienting to arrow cues. These findings are consistent with the idea that reflexive orienting to non-biological cues is underpinned by subcortical brain mechanisms that are shared between the two hemispheres, whereas reflexive orienting to gaze cues is subserved by lateralized cortical mechanisms involved in face/gaze processing (e.g., Kingstone, Tipper, Ristic & Ngan, 2004; Friesen & Kingstone, 2003a).

Selection of gaze cues

While this extant body of research has made a number of major inroads into the shift of attention that is triggered by a gaze cue, it has left relatively untouched the initial selection of gaze information prior to an attention shift to a gazed-at location. Both of the prominent models of social attention claim that the eyes are the primary cue to the direction of social attention
(Baron-Cohen, 1994; Perrett et al., 1992). If this is true, then as social organisms we should be highly sensitive to the presence of eyes (Baron-Cohen, 1994). A result of this sensitivity is that we should be biased to select eyes from our environment. As I will discuss, this critical assumption has been given little empirical consideration.

Before moving on to a literature review of studies of gaze selection, it should be clarified that the selection of gaze is not (and cannot) be measured by the standard gaze cueing paradigm. That is, gaze cueing studies focus exclusively on the shift of attention that occurs after a central directional gaze cue has been processed. In these studies, in order to ensure that cue direction is indeed processed (even though the cue is irrelevant, that is, spatially uninformative), the cue is presented at the current focus of attention, i.e., at the central fixation point, where observers are instructed to look. In addition, when the gaze cue appears there are no other stimuli on the screen competing for selection. In a very real sense, the gaze cue is preselected for the subject by the experimenter. Thus, these studies are concerned solely with the shift of attention that is triggered by the gaze cue, e.g., the time course of this attention shift, its vulnerability to top-down intentions of the participant, and the like. Because these cuing studies only examine the effects of (pre-selected) gaze cues on the spatial orienting component of attention, these studies do not inform us about the selection of gaze cues from the environment.

What evidence is there that humans preferentially select eyes over other stimuli? Several eye movement studies have measured whether adults overtly select (i.e., fixate) eyes more so than other regions of the face. In these ‘face-scanning’ studies, observers routinely spend more time fixating the internal features of the face (eyes, nose, mouth) than the external features (face outline, ears, hair, etc.) (e.g., Althoff & Cohen, 1999; Henderson, Falk, Minut & Dyer, 2001; Walker-Smith, Gale & Findlay, 1977). Perhaps the earliest of these studies was by Yarbus (1967), who showed face portraits to observers and monitored their eye movements while they freely viewed the images for anywhere from 1 to 3 minutes. The scan-patterns from individual subjects showed a preference for the eye and mouth regions of the face, with some tendency for more fixations on the eyes. A later study by Luria and Strauss (1978) found again
that the internal features were fixated more than the external features. Mertens et al. (1993) showed photographs of faces and found that the eyes, mouth, and nose were the preferred targets of gaze. Other studies have found a more explicit preference for eyes over the other facial features. For instance, Walker Smith, Gale and Findlay (1977) asked three subjects to either match a test face with a previously viewed face, or to compare two faces side-by-side. Overall, subjects spent around 70% of their fixations on the eyes. Similarly, a recent study by Pelphrey et al. (2002) found that healthy adults spent approximately 80% of their time fixating the eyes while freely viewing photographs of faces (2% mouth; 10% nose; 8% external features). When asked to identify the emotions in the face, a similar pattern emerged: subjects spent the majority of their time fixating the eyes (55%), less time on the nose (20%), mouth (7%), and external features (15%). Henderson, Williams and Falk (2005) found that observers consistently spent over half of their fixations on the eyes, and even more so (>60%) when trying to remember the face in a later memory test. This is consistent with an earlier study by Henderson et al. (2001), who found that the eyes received approximately 60% of fixation time. Indeed, this preference for the eyes of a face is also evident early on in life. For instance, infants have been found to preferentially scan a face’s eyes by 2 months of age (Haith, Bergman & Moore, 1977; Maurer & Salapatek, 1976). Before this age, infants spend most of their time looking away from the face (Maurer & Salapatek, 1976) or, when looking at the face, at its periphery (e.g., chin and hairline, Maurer & Salapatek, 1976; Haith, Bergman & Moore, 1977).

Given these findings from face scanning studies, one might conclude that eyes are preferentially selected, consistent with the social attention models. However, it is notable that, like the gaze cueing studies of attention, these ‘face scanning’ studies have routinely preselected the face for the observer. That is, the typical procedure is to present a photograph of a single face (black-and-white or coloured) at the centre of the screen, isolated from its body parts and surrounding context. Although the face scanning paradigm differs from the gaze cueing paradigm in that observers are free to move their eyes over the display, i.e., they are free to select specific facial features for further processing, preselecting the face may lead to
uncertainty in interpreting why exactly observers are biased to look at the eyes. One concern is that because the face is presented in isolation, observers are limited to selecting among the few items within the face (e.g., eyes, nose, mouth, chin). Furthermore, when a face is presented in this ‘passport’ style, the eyes may be the most visually salient and interesting part of the display (e.g., see Figure 1.1). For instance, the eyes are notably high contrast features of the face, with a dark iris and white sclera, and they are surrounded by (usually) dark brows that frame the upper portion of the eye socket. Thus, it may be the case that observers select the eyes not because they are a socially important, but simply because they are more visually salient or interesting than the other facial features available for selection (e.g., Neumann et al., 2005; Itti, Koch & Niebur, 1998). If this is true, it would present a problem for the notion that humans have a special sensitivity to the eyes because they provide important social information.

This issue deserves further elaboration. The fact that the eyes are visually salient is not a problem per se. Indeed, a feature that distinguishes human eyes from the eyes of other primates is the high contrast of the iris and pupil relative to the white sclera (Kobayashi & Kohshima, 1997; Langton et al., 2000). This evolutionary change in the eye stimulus is argued to have occurred so that humans could read more complex social signals from the face – one of these social signals being the direction of gaze of others (Kobayashi & Kohshima, 1997). In addition, it has been suggested that infants begin to preferentially scan the eyes of adults precisely because they are visually salient (Maurer & Salapatek, 1976), becoming more so as the infant’s visual acuity improves. Thus, it may be that the saliency of eyes initially attracts the attention system precisely because eyes are an important social communication signal. With time, the infant learns that the eyes of others are useful social communicative cues, providing valuable information about the mental states of others. Whether or not the eyes become preferential social stimuli because of their visual salience, or not, is an open question. However, despite the reasons for the development of a preference for eyes, it stands to reason that visual salience is not what drives a fully developed human to seek out eye stimuli. The models of social attention propose that we preferentially select eyes from our environment because they
provide important social-communicative information, not because they are visually salient (e.g., Baron-Cohen, 1995). It is this crucial assumption that needs to be tested more rigorously.

All of the face-scanning studies suffer from the potential fundamental flaw that the eyes may have been preferentially selected due to their visual conspicuity alone. Thus, there is at best only equivocal support for the notion that humans have a special preference for eyes relative to other facial features, above and beyond what might be predicted by bottom-up models of attentional selection (e.g., Itti, Koch & Niebur, 1998). Because the preferential selection of eyes in face scanning studies can be explained by either the social importance of eyes or the visual saliency of eyes, this renders face scanning studies inadequate for assessing the claim that observers have a natural bias to select eyes because they are special social stimuli. That is, while the preselection of gaze stimuli is perhaps appropriate for studying post-selective processes such as spatial orienting to gazed-at locations, it is certainly not appropriate for studying whether eyes are preferentially selected over other stimuli. In order to test this, eyes should be presented among other non-eye stimuli of equal (or greater) visual salience that might compete for attention.

What do people look at when the eyes are not the most salient items in a scene? The early work of Yarbus (1967) sheds some light on this issue. In addition to an isolated face (Figure 1.2a), participants were shown a few images of an entire body or of the face and part of the person’s trunk (Figure 1.2b). The scan patterns for these images show widely distributed fixations, with several clustering on the face. However, there was no obvious preference for eyes in these scan patterns. In addition, unfortunately, the resolution and presentation of Yarbus’ eye tracking data do not allow for a clear picture of where observers actually fixated (e.g., see Figure 1.2b). These results point to the possibility that when the eyes are presented in a face along with other body parts, and are possibly no longer the most visually salient stimuli, they are no longer preferentially selected. This is disturbing for the wide-spread assumption that humans preferentially select eyes over other stimuli.

Conflicting evidence, however, comes from another dataset in Yarbus’ study. Yarbus showed participants a picture of Repin’s painting, "An Unexpected Visitor" and found that there
was a tendency to look at the heads and faces of the people in the scene (see Figure 1.2c). This discrepancy might suggest that increasing social content of a scene, by adding more people to it, may increase the tendency for observers to look at the eyes of other people. However, again, the resolution of Yarbus' eye monitor does not allow the reader to distinguish between fixations that observers made to the face, eyes, or other regions of the heads of the people in the painting.

Perhaps the only noteworthy evidence for a preferential selection of eyes comes from a recent study by Klin et al. (2002). Klin et al. examined eye movements while observers (typically developing adults and individuals with autism) watched dynamic social scenes. Five clips, lasting 30-60 seconds long, were taken from the movie, “Who’s afraid of Virginia Woolf?”. The eye movement data showed that typically developing observers spent on average 65% of their fixations on the eyes of the actors, 21% on mouths, 10% on the bodies, and 4% on objects (although there were very few focal objects in the scenes). Viewers with autism, on the other hand, spent more fixations on the mouths (41%) than the eyes (25%), bodies (25%) or objects (10%). Thus, Klin et al. found evidence that healthy adults do preferentially select the eyes over other stimuli. What is particularly favourable about their approach is that observers were shown the eyes among complex social scenes, containing many other salient objects and scene regions available for selection. However, several concerns exist surrounding their methodology. The most serious of these is that their coding scheme may have biased the fixation proportions. In their paper, Klin et al. first describe how they divided the whole on-screen area into three regions: face, body, objects. Later, however, they mention that they further subdivided the face region into two regions of interest: eyes and mouth. Thus, in the end, their coders were asked to code whether a fixation landed on eyes, mouth, bodies, or objects. This coding scheme means one of two possible things: either the coders discarded fixations that landed on a part of the face other than the eyes and mouth (e.g., nose, hair, cheek, ear), or the face was actually divided along the horizontal meridian, with anything above this meridian coded as “eyes” and anything below this meridian coded as “mouth”. It is unclear from their paper which of these two options is true. Both are problematic, however, and may have significantly biased the fixation
proportions for the eyes and mouths. In addition, Klin et al. do not make it clear what they considered to be “objects”. Did any fixation that fell outside of a person count as an “object” (e.g., walls, floor, couch, etc.), or did the fixation have to land on a foreground object (e.g., cup on table)? These details are critical to the validity of the conclusion that healthy observers spent 65% of their fixations on the eyes, because if many fixations were discarded because they did not count as landing on eyes, mouth, or objects, then the reader is not receiving a clear picture of subjects’ overall fixation preferences. Therefore the empirical support for the special status of eyes relative to other biological social items (e.g., body, arms, legs) and non-social items is less clear-cut than one might initially think.

Thesis overview

The general goal of the thesis is to examine whether gaze is preferentially selected from complex visual scenes; and if selection is biased towards eyes, then why? The majority of research on social attention has failed to address the fundamental issue of whether people really do preferentially select eyes from complex scenes. This is the case both because gaze selection has rarely been examined empirically and because, on the rare occasions it has been tested, the data do not provide any clear support for the idea that eyes are prioritized within complex scenes (e.g., Yarbus, 1967; Klin et al., 2002).

The thesis will address five main research questions:

(1) Is there a preferential bias to select eyes from complex social scenes?
(2) What factors influence this selection process?
(3) To what extent does the preferential bias to select eyes generalize to different tasks and situations?
(4) What is the role of visual saliency in driving fixations to the eyes within complex social scenes?
(5) How does studying the selection of social cues shed light on past controversies in the social attention literature?
Chapter 2 presents the initial investigation of selection within complex social scenes. The findings show a preferential bias to select eyes relative to other scene regions. In addition, this study shows that the bias to select eyes is enhanced by social factors, specifically by increasing the social content and activity in a scene. Chapter 3 examines additional effects of task instruction, showing that a task demanding participants to report on the social attention of people in the scene enhances fixations to the eyes relative to less social tasks. Chapter 4 asks whether there is a default bias to select the eyes across all tasks or whether this only occurs in tasks encouraging observers to attend to the people in the scene. The results show a default bias to preferentially select the eyes across six different tasks, including the instruction to study the uninformative aspects of the scene. However, the preference for eyes was enhanced by instructions to study the mental states of people in the scenes. Chapter 5 asks whether observers who are told to remember scenes perceive the eyes to be important for this task. The results show that observers who were told about a later memory test on the scenes enhanced their fixations to the eyes relative to observers who were not aware of this memory test. Furthermore, when recognizing the scenes, observers tended to focus on the same regions as they did when the scenes were first viewed. Chapter 6 assesses the role of visual saliency in determining where observers look within complex social scenes. The results indicate that saliency does not account for observers’ interest in the eyes, nor does it explain a significant amount of fixation behaviour. Chapter 7 examines selection of eyes compared to another directional and socially communicative cue, the arrow. Gaze-cueing and arrow-cueing studies of spatial orienting are discussed, and it is noted that these studies have largely failed to capture the unique social importance of eyes. In the General Discussion (Chapter 8) the implications of the collective data are discussed, along with future directions.
Figure 1.1. The salient regions of a face. Yellow lines outline salient regions computed by Itti & Koch’s (2000) saliency model. Red lines indicate predicted path of attention among the salient regions.
Figure 1.2

I was unable to obtain permission from the copyright holder to reproduce this figure in my dissertation electronically. Please visit http://www.informaworld.com/smpp/content~content=a780895549~db=all to see the figure electronically.

Figure 1.2. Scanpaths produced in Yarbus’ (1967) study, in which participants freely viewed a set of images. A. Scanpaths for an individual face, showing a selective preference for the eyes. B. Scanpaths for a face accompanied by the rest of the body. Note that the preference for eyes is greatly reduced compared to when viewing an individual face. C. Scanpaths for a social scene, Repin’s An Unexpected Visitor. Here a preference for faces (and possibly eyes) is again observed. From Eye Movements and Vision, by A. L. Yarbus, 1965 (translated by B. Haigh, 1967), New York: Plenum Press. Copyright 1965 by Springer Science and Business Media (pp. 174, 180, and 189).
References


CHAPTER 2

Social Attention and real world scenes:

the roles of action, competition, and social content.

Over the last decade there has been an explosion of research interest in what has become known as “social attention”. This research has generally focused on understanding how one's attention is affected by the presence of other individuals, epitomized by studies showing that infants and adults alike will attend automatically to where someone else is looking (Hood, Willen & Driver, 1998; Friesen & Kingstone, 1998; see Langton, Watt, and Bruce, 2000 for a review). In the typical laboratory investigation of this sort, a participant is first shown a picture of a real or schematic face with the eyes looking sideways toward the left or right. Shortly thereafter a response target is presented either at the gazed-at location or the nongazed-at location. One normally finds that response time (RT) to detect a target is shorter when the target appears at the gazed-at location than the nongazed-at location. Importantly, this effect emerges rapidly and occurs even when gaze direction does not predict where a target is going to appear.

Although these data fit nicely with the intuition that attention is shifted to where people are looking because one cares about where other people are attending, recent research suggests that this original interpretation (e.g., Friesen & Kingstone, 1998; Langton et al., 2000) may have overstated its case. It has been shown that other biological social-communicative cues produce an attention effect that is very similar (Friesen, Ristic, Kingstone, 2004), if not identical (Ristic et al., 2002; Tipples, 2002), to what is found for gaze direction. Some of these cues, like head direction (Langton, 2000) and pointing fingers (Watanabe, 2002), obviously pertain to people, and thus their effects can be readily accommodated by the notion that a range of biological social cues that indicate where other people are attending will trigger a shift in one's attention (Kingstone et al., 2003). However, it is now clear that almost any cue with a directional component, from arrows (Ristic & Kingstone, 2005; Tipples, 2002) to numbers (Fischer et al., 2003; Ristic, Wright & Kingstone, 2006) will produce a shift in one's attention. These latter cues are not directly associated with people, suggesting that biological social-communicative stimuli may not hold any special status when it comes to producing shifts of attention in cuing studies. What is critical is simply that the cues are directional in nature.
Does it follow from this conclusion that biological social-communicative cues, like the eyes, are never given preferential status by the attentional system? No. After all, there is a wealth of data in the research literature indicating that when a picture of a face is presented to participants, they look preferentially (70-80% of the time) at the eyes of that face (Pelphrey et al., 2002; Henderson et al. 2000; Walker-Smith et al., 1977). This would suggest that the attentional system has a preferential bias for the eyes. And yet, if we have learned anything at all from the cuing studies reviewed above, it is that eyes cannot be considered "special" unless one has compared them against non-face stimuli.

With this point in mind, it is sobering to note that the vast majority of the face scanning studies have mainly presented participants with only a face to look at. Indeed, not only are all other kinds of background information routinely stripped away, but faces normally are presented in a 'passport-type' format, with all body features below the neck missing. Thus, it is possible that people look preferentially at the eyes of an isolated face simply because eyes are the most interesting stimuli in a relatively impoverished display. Therefore the special status of eyes relative to other biological social items (e.g., body, arms, legs) and non-social items is less clear-cut than one might initially think.

Indeed, looking at the literature with this thought in mind, one cannot help but be struck by how little research data there are concerning how people look at scenes when a face is observed as it is normally perceived in the real world, that is, among other body parts, faces, and nonsocial items. On this score past investigations are remarkably silent, save for one most noteworthy exception – the seminal work by the Russian physiologist Alfred Yarbus (1967). Yarbus recorded the eye movements of subjects looking at pictures, continuing the earlier work of Buswell (1935). Yarbus found that when observers are shown the picture of a face presented in isolation, observers tend to look at the eyes of the face, whether it is human or another animal (see Figure 2.1a). As noted above, this preference for the eyes of an isolated face is now a well established finding. What has been overlooked, however, is that Yarbus also found that if the face is not presented in isolation, but accompanied by its associated body parts such as arms,
torso, and legs, the preferential scanning of the eyes tends to disappear both for humans and other animals (Figure 2.1b). Thus Yarbus’ data suggest that there may not be a preference for scanning the eyes when the face of a person is presented along with its body. Following a similar line of reasoning, the preference for eyes might also be expected to decline if there were other objects competing for attention, such as those typically found within a complex scene. The present investigation put precisely this hypothesis to the test.

---Figure 2.1---

Competition from other objects may not, however, be the entire story. We noticed that when Yarbus showed his participants a painting by Repin that depicted several people in a room (see Figure 2.1c), participants now tended to look at the faces of the people in the scene. This suggested to us that increasing the social content of a scene, by adding more people to it, may increase participants’ interest in the eyes of the characters. Unfortunately, the resolution of the Yarbus’ eye monitor did not discriminate clearly between observers’ scanning of the eyes in the scene from the other available facial features. And most critically, there is the concern that all of the observers tested by Yarbus were well acquainted with Repin’s picture, a concern raised by Yarbus himself: “This evidently accounts for the generally considerable similarity between all the records. … Undoubtedly, observers familiar both with the picture and the epoch represented in it would examine the picture differently from people seeing it for the first time and unfamiliar with the epoch it represents.” [Yarbus, 1967, p.192]. This concern is reinforced by Yarbus’ subsequent demonstration that if people are simply asked different questions about the picture, their fixation pattern can vary dramatically and systematically from the free-viewing conditions. In other words, the fixations on the heads and eyes in the Repin picture may have reflected a shared knowledge of the picture being viewed and the problem it posed for the knowledgeable

---

1 We overlaid the scan paths from Yarbus (1967) onto their corresponding pictures. However, the scanpath image was often not the same size as the picture image, and so an exact match was not possible. As a result, it is actually very difficult to know where the clusters of fixations landed within Yarbus’ original images, again making it unclear whether eyes received preferential scanning in all (or any) of his pictures.
observer at the time of perception. These considerations, coupled with the fact that Yarbus’ observations stem from the use of a single visual scene, leads one to the conclusion that there is a need to examine further the influence of social content on how attention is allocated, with special interest paid to whether it impacts the allocation of attention to the eyes.

In sum, the present study had two main goals. First, we wanted to determine whether, with naïve observers, in free-viewing conditions, and with other items readily available for viewing, there is a preference for scanning the eyes of a single individual in a scene. Second, we wanted to discover whether adding more people to a scene will increase the degree that eyes are scanned.

It is worth noting that, when manipulating the number of people within different scenes, one is immediately faced with the problem of what the people in the scenes should be doing. Because we had no way of knowing how action in a scene would impact scanning patterns (e.g., Repin’s painting is ambiguous on this score) we controlled and manipulated this factor by having people photographed either doing nothing (e.g., just sitting on their own; Inactive scenes), doing something (e.g., sitting reading a book on their own; Active scenes) or, in the case when there were multiple people in a scene, doing something separately (e.g., sitting together but reading individually; Active Scenes) or doing something together (e.g., sharing a book; Interactive scenes). These latter conditions, in which multiple people in a room are either acting separately or interacting, represent a subtle yet potentially important difference in social content, and thus in keeping with the aims of the present study we wished to examine whether the eyes would be scanned differently in scenes containing social action (several people doing something separately) compared to scenes with social inter-action (several people doing something together). Examples of these scene-types are presented in Figure 2.2. Specific experimental details are presented below.

--Figure 2.2--
Method

Participants

Twenty undergraduate students from the University of British Columbia participated in this experiment. All had normal or corrected to normal vision, and were naïve to the purpose of the experiment. Each participant received course credit for participation in a one-hour session.

Apparatus

Eye movements were monitored using an Eyelink II tracking system. The on-line saccade detector of the eye tracker was set to detect saccades with an amplitude of at least 0.5°, using an acceleration threshold of 9500°/s^2 and a velocity threshold of 30°/s.

Stimuli

Full color images were taken with a digital camera in different rooms in the Psychology building. Image size was 36.5 x 27.5 (cm) corresponding to 40.1° x 30.8° at the viewing distance of 50 cm, and image resolution was 800 x 600 pixels. Forty scenes were used in the present experiment. Each scene contained either 1 or 3 persons. In the 1-person scenes the individual was either doing something (Active) or nothing (Inactive). Similarly, in the 3-people scenes, people either did nothing (Inactive), or did something on their own (Active), or did something together (Interactive). All scenes were comparable in terms of their basic layout: each room had a table, chairs, objects, and background items (e.g., see Figure 2.2).

Procedure

Participants were seated in a brightly lit room, and were placed in a chin rest so that they sat approximately 50 cm from the display computer screen. Participants were told that they would be shown several images, each one appearing for 15 seconds, and that they were to simply look at these images.

Before beginning the experiment, a calibration procedure was conducted. Participants were instructed to fixate a central black dot, and to follow this dot as it appeared randomly at nine different places on the screen. This calibration was then validated, a procedure that calculates the difference between the calibrated gaze position and target position and corrects
for this error in future gaze position computations. After successful calibration and validation, the scene trials began.

At the beginning of each trial, a fixation point was displayed in the centre of the computer screen in order to correct for drift in gaze position. Participants were instructed to fixate this point and then press the spacebar to start a trial. One of 40 pictures was then shown in the center of the screen. Each picture was chosen at random and without replacement. The picture remained visible until 15 seconds had passed, after which the picture was replaced with the drift correction screen. This process repeated until all pictures had been viewed.

Results

Data handling

For each image, an outline was drawn around each region of interest (e.g., "eyes") and each region’s pixel coordinates and area were recorded. We defined the following regions in this manner: eyes, heads (excluding eyes), body (including arms, torso and legs), foreground objects (e.g., tables, chairs, objects on the table) and background objects (e.g., walls, shelves, items on the walls). Figure 2.2 illustrates these regions.

To determine what regions were of most interest to observers we computed fixation proportions by dividing the number of fixations for a region by the total number of fixations over the whole display. These data were area-normalized by dividing the proportion score for each region by its area (Smilek, Birmingham, Cameron, Bischof & Kingstone, 2006).

To determine if observers’ interest in the regions changed over time we computed the cumulative fixation proportions for the regions in 1-second time steps for the 15 seconds of display duration. These data were area-normalized by dividing the proportion score for each region by its area.

To determine whether fixation time differed among the regions we computed the duration proportions for each region. These data were area-normalized by dividing the time score for each region by its area.
Fixation proportions

Before analyzing all the data we examined whether participants viewed the “3-people Active” scenes differently from the “3-people Interactive” scenes. These two scene types were scanned similarly (i.e., there were no significant differences in fixation proportions between comparable regions, all Fs<1), so we combined the data from these two scene types to create a single data set of “3-people Active”. We then submitted all the fixation proportion data to a 2x2x5 within-subjects analysis of variance (ANOVA) with People (1 person vs. 3 people), Activity (Inactive vs. Active) and Region (Eyes, Head, Body, Foreground, Background) as factors.

Figure 2.3 shows these data for eyes, heads, and other regions. Looking at this figure it is immediately evident that the eyes were fixated far more than any other region, as reflected by a main effect of Region, F(4,76)=639.86, p<0.001. This strong preference for the eyes was true for both 1-person scenes (0.60 average fixation proportion) and 3-person scenes (0.64 average fixation proportion). Furthermore, it is clear that while eyes received more fixations than all other regions, heads were also fixated quite frequently, more so than bodies, foreground objects, and background. In addition, there were no differences between bodies, foreground objects, and background. These observations were confirmed with post hoc Tukey-Kramer pairwise comparisons (p<0.05).

A significant Region x People x Activity interaction (F(4,76)=14.88 p<0.001) reflected, however, that when there was activity in the scene, there were more fixations on the eyes when 3 people were in a scene than when 1 person was in the scene. This observation was confirmed by a post hoc pairwise comparison (Tukey-Kramer p<0.05) of 3-people (0.66 fixation proportions) versus 1-person Active scenes (0.56 fixation proportions). In contrast, when there was no activity in a scene, there were no more fixations on the eyes for 3 people scenes than 1 person scenes (Tukey-Kramer p>.05). Further pairwise comparisons (p<0.05) revealed that observers fixated the eyes more in the 3-Active scenes that in the 3-Inactive scenes, whereas
the opposite effect of activity occurred for one person scenes (1-person Inactive > 1-person Active).²

**Cumulative fixation proportions**

We were also interested in how scanning preferences changed over time. An ANOVA was performed on the cumulative fixation proportion data with People, Region, Activity and Interval as factors, the latter broken into one-second intervals (0-1s, 1-2s,...,14-15s). Figure 2.4 shows these data for eyes, heads, and other regions. All of the effects from the fixation proportion analysis were again significant in the analysis of cumulative probabilities. For example, there was a main effect of region, reflecting the fact that again eyes were fixated by far the most, $F(4,7200)=12406.46$, $p<0.001$ (posthoc comparisons revealed that this preference for eyes was significant ($p<0.05$) even at the first one-second interval)³. Looking at Figure 2.4 one sees that many of the regions were re-fixated over time, resulting in an overall increase in the cumulative fixation proportions, and hence a main effect of Interval, $F(14,7200)=483.58$, $p<0.0001$. This increase in cumulative fixation proportions over time was most pronounced for eyes, reflected by a Region x Interval interaction, $F(56,7200)= 153.55$, $p<0.0001$.

---Figure 2.4---

Recall that the key finding from the overall fixation proportion data was that there was an overall preference for the eyes in 1- and 3-person scenes, with scene activity enhancing the preference for eyes in the 3-people scenes. Our cumulative analysis revealed that this preference for eyes in the Active 3-people scenes did not begin to emerge until after 6 seconds, at which point there was a significant difference between fixations on the eyes in 3-people

---

² We were concerned that the overall interaction between people and activity and the bias it has on scanning the eye region might be an artifact of averaging across a few specific actions. Importantly, however, across 8 of the 9 active scenes (ranging from simply reading a book to threatening to punch each other) there was a large preference for the eyes in the active 3-people scenes compared to the active 1 person scenes, with a binomial test revealing that the probability of 8 or more rooms showing this magnitude difference being $p< 0.02$.

³ While eyes were fixated the most within the first 1-second interval, the very first fixation was made most often to the head than any other region, $F(4,76)=24.56,p<0.001$. This movement to the head would appear to be a “signature” of first acquiring the eye region.
Active (yellow circles) relative to 1-person Active scenes (white circles). This difference persisted until the end of viewing. This observation was confirmed by a People x Activity x Interval interaction for cumulative fixations on eyes, and subsequent post hoc pairwise comparisons (p<0.05).

**Duration proportions**

Duration proportions are shown in Figure 2.5. These data closely parallel the fixation proportion data. For instance, it is clear that the eyes were fixated for the longest out of all the regions, followed by heads, bodies, background, and foreground objects. This was reflected in a highly significant main effect of Region (F(4,76)= 552.23; p<0.0001). Post hoc pairwise comparisons (p<0.05) revealed significantly longer durations for eyes than for heads, bodies, background, and foreground objects. In addition, heads were fixated for longer than bodies, background, and foreground objects (p<0.05). In addition to the effect of Region, there was also a People x Activity x Region interaction (F(4,76)=17.42, p<0.0001). This latter, higher order interaction indicated a similar pattern as in the fixation proportion analysis. That is, when there was activity in the scene, fixation time was longer on the eyes when 3 people were in a scene than when 1 person was in the scene. This observation was confirmed by a post hoc pairwise comparison (p<0.05). In contrast, when there was no activity in a scene, fixation durations were equal for 3 people scenes and 1 person scenes, p>0.05. As in the fixation proportion analysis, further pairwise comparisons (p<0.05) revealed that observers fixated the eyes longer in the 3-Active scenes than in the 3-Inactive scenes, whereas the opposite effect of activity occurred for one person scenes (1-person Inactive > 1-person Active).

---Figure 2.5---

**Discussion**

The present study had two main goals. First, we wanted to determine whether, with naïve observers, in free-viewing conditions, and with other items readily available for viewing, there is a preference for scanning the eyes of a single individual in a scene. While it was well-established in the literature that people look at the eyes of a face when the face is the only item
available for scanning, it was not at all clear that this finding would hold when a person was presented with a face along with its body and other items in a scene. Indeed we had noted that the seminal work of Yarbus provided evidence to suggest that scanning of the eyes might not receive preference relative to other body parts and/or objects if they were made available for viewing. The results of our study were unequivocal on this issue. People prefer to look at the eyes of one person in a scene, even when there are other items available (average fixation proportions 0.60 (eyes) versus 0.40 (elsewhere)). Thus, the preferential bias for the eyes of a person persists in real world scenes containing other body parts and objects. However, it is also clear that eyes did not entirely dominate observers’ attention in these scenes, as fixations did frequent (.40) other body parts and objects in the scenes. For instance, the region with the next highest fixation proportion, and duration proportion, was the head region, which was fixated more frequently and for longer than the body region, foreground objects, and background objects (but less so than eyes). Thus, it is clear that observers showed a particular interest in the faces of people in the scenes, focusing especially on their eyes.

The present study also demonstrated that the preferential bias for eyes emerged remarkably early. Our analysis of the cumulative fixation proportions showed that after first fixating the head, the eyes received more fixations than other regions within the first second of viewing. Moreover, observers were more likely to revisit the eye region again and again while viewing a scene, resulting in an enhanced preference for eyes as viewing time was extended.

The second main goal of our study was to discover whether increasing the social content of a scene, by adding more people to it, would increase the extent that the eyes are scanned. We had noted that the eye-scanning data of the Repin painting (Yarbus, 1967) provided indirect support for our proposal that people will look more to the eyes as the social content of a scene is increased. Convergent with this proposal is the recent finding that people look more to the eyes as the need to extract the social information of a scene increases, for instance, in order to infer the attentional states of the people depicted within a scene (Smilek et al., 2006).
The data from the present study shed new light on this issue. First, we found that increasing the social content of a scene does drive people to look more, and longer, at the eyes, but only when the people in the scene are actively doing something. Second, our cumulative probability data revealed that this impact of social content and activity does not first emerge until after 6 seconds of viewing time, suggesting that this interaction reflects a rather complex level of scene-analysis by the observer. It is our speculation that when the social content of a scene is relatively low, that is, there is only one person in the scene, action draws attention away from the eyes because eye information is not critical to understanding the action. However, when the social content of a scene is relatively high, that is, there are multiple people within a scene, action draws attention toward the eyes because eye information is critical to understanding the social meaning of the action.

It is clear that this speculation requires future investigation. Nevertheless, it is noteworthy that our data do suggest a subtle, yet powerful, way to examine the observers’ sensitivity to changes in social content. For instance, it would be interesting to examine how people with autism scan scenes with one versus many people; and how their exploration of these scenes is affected by the action within it. Previous studies have shown that individuals with autism are less likely to spontaneously orient to social stimuli (e.g., people) in their natural environments, (Dawson et al., 1998; Osterling & Dawson, 1994; Osterling, Dawson, & Munson, 2002; Swettenham et al., 1998) and demonstrate abnormal face processing compared to typically developing individuals (e.g., Behrmann et al., 2006; Langdell, 1978; Joseph & Tanaka, 2002). Eye movement studies have shown that these social impairments are further characterized by a specific avoidance of eyes (Klin et al., 2002; Pelphrey et al. 2002). One explanation for these findings is that individuals with autism have a heightened negative emotional response to the eyes of others, and that they avoid the eyes in order to reduce this over-arousal (Dalton et al., 2005). This hypothesis would predict that individuals with autism are averse to an increase in social content. If so, then they might tend to look away from the eyes as people are added to a scene and their activity increases.
An intriguing possibility is that because eye information may be critical to understanding the social meaning of action when there are three people in a scene, observers might also be likely to make more eye movement transitions between the eyes of the people in 3-people Active scenes. That is, perhaps in trying to understand the nature of the social activity occurring in 3-Active scenes, observers would have made more eye movements between the eyes of the different people in the scene to look for states of shared or mutual attention. If this were true, and observers made more eyes-eyes transitions in these scenes relative to 3-people Inactive scenes, it would bolster our conclusion that eye information is critical to understanding the meaning of social action. However, an analysis of eyes-eyes transition frequencies revealed no such differences (p>0.05). In fact, if anything, there was a nonsignificant trend toward fewer eyes-eyes transitions for 3-Active scenes (M=0.17) than 3-Inactive scenes (M=0.23). Thus, it may be the case that fixation frequencies and fixation durations are more sensitive to the effects of social content and activity that are transition frequencies. Alternatively, it may be that differences in transition frequencies occurred, but that they were masked by the variation in social activity occurring across the scenes (e.g., variations in mutual versus shared gaze, etc.). Future studies will be required to investigate this possibility further.

Collectively the data from the present study provide support for our interpretation of Yarbus’ original work, from which we had hypothesized that scanning of the eyes would be sensitive to the presentation of competing objects and variation in social context. Our study goes well beyond this initial work, however, in systematically testing and confirming that these are important factors to consider when studying how people scan natural scenes. Moreover our work has implications for the social attention literature. Whereas several studies have shown a preferential bias for eyes within highly simplistic displays (e.g., cutout of a face against a blank background), to our knowledge the present study provides the first demonstration that this bias is expressed for complex real world scenes. Our study also provides intriguing evidence that as

\[4\] We thank an anonymous reviewer for this suggestion.
scene complexity continues to increase, e.g., by adding action and social content to a scene, the preference for the eyes will continue to be enhanced.

Finally, our work has implications for recent scene scanning studies, insofar as we drive home the fact that the social content of a scene is critical to where people look within a scene. To date most studies of this nature have used scenes that do not contain people (de Graef, 1992; Henderson & Hollingworth, 1999; Hollingworth, Weeks & Henderson, 1999; Rayner & Pollatsek, 1992; Underwood & Fousham, 2006). Thus, an important question is whether the results from these previous studies are restricted to scenes without people. Our present work suggests that they may well be, i.e., people prefer to look at people (for support for this idea see the work of Fletcher-Watson, Findlay, Leekam & Benson, 2008). It is important to note that we are not claiming that this preference for people – especially the eyes of people – is absolute. It is well established that eye movements change according to task demands. That said, it remains an open question how changes in task demands will affect people’s profound tendency to look at the eyes of others.
Figure 2.1

I was unable to obtain permission from the copyright holder to reproduce this figure in my dissertation electronically. Please visit http://www.informaworld.com/smpp/content~content=a780895549~db=all to see the figure electronically.

Figure 2.1. Scanpaths of an individual face (A), a face accompanied by the rest of the body (B), and a social scene (C). From Eye Movements and Vision, by A. L. Yarbus, 1965 (translated by B. Haigh, 1967), New York: Plenum Press. Copyright 1965 by Springer Science and Business Media (pp. 174, 180, and 189).
Figure 2.2. A. Examples of the four scene types. From top to bottom: 1-person Active, 1-person Inactive, 3-people Active, 3-people Inactive. B. Corresponding regions of interest used in analysis (eyes, head, body, foreground objects, background objects). C. Corresponding plots of fixations for all subjects.
Figure 2.3. Fixation proportion data plotted as a function of People, Activity, and Region. Fixations to eyes were enhanced by increasing social content (i.e., 3 people scenes vs. 1 person scenes) when the scenes contained activity (Active scenes).
Figure 2.4. Cumulative fixation proportions for 1-person scenes and 3-people scenes. Data are plotted as a function of region, activity level, and viewing interval.
Figure 2.5. Duration proportion data plotted as a function of People, Activity, and Region. Fixation time on eyes was enhanced by increasing social content (i.e., 3 people scenes vs. 1 person scenes) when the scenes contained activity (Active scenes).


CHAPTER 3

Gaze selection in complex social scenes

Imagine the following scenario. You are walking down a busy city street and you notice that there is a woman who has stopped walking and is gazing upward. Using her gaze direction you turn your eyes to see what she is looking at.

As this simple example illustrates, folk knowledge suggests that we are very interested in the attention of other people, and that we use their eyes to infer where, and what they are looking at. Moreover, we seem to do this in at least two distinct stages. First, we select the eyes as a key social stimulus, and second, we shift our attention from the eyes of someone to the location/object that someone is looking at.

To date research has focused on the second stage of this equation by examining the extent to which gaze direction can trigger an attention shift in others and seeking to uncover the neural circuitry that subserves this attention shift. As a result it is now firmly established that infants (Hood, Willen & Driver, 1998), preschool children (Ristic, Friesen & Kingstone, 2002) and adults alike (Driver et al., 1999; Friesen & Kingstone, 1998; Langton & Bruce, 1999) shift attention automatically to where others are looking. Single cell (Perret et al., 1985), brain lesion (Campbell et al., 1990), and functional neuroimaging studies (Bentin et al., 2002; Dolan et al., 1997; Kingstone et al., 2004) have implicated specific brain areas, such as the superior temporal sulcus and the superior parietal lobe, as critical neural components of this orienting process.

While this extant body of research has made a number of major inroads into the shift of attention that is triggered by a gaze cue, it has left relatively untouched the question of what factors are critical to the initial selection of gaze information prior to an attentional shift to a gazed-at location. Indeed, in a typical study the social cue -- that is, the eyes -- is preselected by the experimenter so that it, along with its associated facial features (e.g., head, nose, mouth), are the only stimuli that a participant receives. As a result, in a typical study participants are presented just with a face (often a schematic face) with no other body parts or objects shown to the subjects. Clearly this approach of preselecting and isolating the social cue circumvents the
critical issue as to what factors are important to the selection of gaze information when it is embedded in complex real world situations\(^1\).

This point would perhaps be moot if it were firmly established that gaze is normally selected in complex visual scenes. However, this matter is far from confirmed, both because it has rarely been examined empirically and because, on the rare occasions it has been tested, the data do not provide any clear support for the idea that eyes are prioritized within complex scenes. The seminal work of Yarbus (1967) provides an ideal demonstration of these points.

Continuing the earlier work of Buswell (1935), Yarbus is one of the few investigators to have examined how people scan scenes containing complex social information. To be sure, Yarbus is well-known for showing that people will look preferentially at the eyes of a face that is displayed in isolation (see for example Figure 3.1a). But what has routinely been overlooked is that Yarbus also studied how people examine images that contain a face along with its associated body parts. An examination of these data, an example of which is illustrated in Figure 3.1b, reveals that there is no obvious preferential scanning of the eyes relative to other parts of the body. These data raise the possibility that the selection of gaze information may not have priority within many complex real world situations.

--Figure 3.1--

Interestingly, in another well-known study, Yarbus showed participants a picture of the Repin painting "An Unexpected Visitor" and found that there was a tendency to look at the heads and faces of the people in the scene (see Figure 3.1c). Note that this finding conflicts with Yarbus’ data above showing that the eyes are not prioritized when a lone individual is

\(^1\) It is worth noting that the routine preselection of gaze information may also have led researchers to over-estimate the importance of gaze direction on the orienting of attention, at least within standard research paradigms. For instance, it has recently transpired that other directional cues, such as arrows (Ristic, Friesen & Kingstone, 2002; Tipples, 2002) and the words "left" or "right" (Hommel et al., 2001), can produce rapid, reflexive shifts of attention that closely approximate (if not duplicate) the attention shift triggered by gaze. This raises the real possibility that many of the orienting effects to gaze direction that were initially attributed to gaze being a "special social cue" (e.g., Friesen & Kingstone, 1998; Driver et al., 1999) may have grossly overstated their case (Ristic, Wright & Kingstone, 2006; Gibson & Kingstone, 2006, for further considerations of this matter).
viewed along with his/her associated body parts. This discrepancy suggested to us two possible explanations. One is that increasing the social content of a scene, by adding more people to it, may increase the tendency for observers to look at the eyes of other people. An alternative explanation is that it is not the social content of the scene per se that is critical, but the level of activity within it, that enhances fixations to the eyes. For instance, in Repin’s painting the characters were all doing something (e.g., walking, opening the door), and the eyes may have contained information that was important for interpreting these activities.

A concern for both of these proposals, however, is that the participants in Yarbus’ study were familiar with Repin's painting and its meaning, and that their scanning of this picture may have reflected a shared understanding of the painting. In other words, the "task-set" that participants brought to the situation may have impacted the scanning of the scene itself. Importantly, Yarbus himself raised precisely this caution regarding his Repin study, and reinforced this consideration by demonstrating that he could change people's scanning patterns simply by asking them different questions regarding the picture. For instance, if he asked observers to remember the clothes worn by the people in the painting, then the observers did not focus on the heads and eyes any longer but on the clothes the being worn by the people in the painting. There are also, unfortunately, a number of other shortcomings related to the Repin study. Besides testing only a small number of participants with a picture that they were intimately familiar with, it was the only picture Yarbus presented to the participants. Thus one does not know whether Yarbus' findings are particular to the situation depicted in Repin's painting, or whether Yarbus' findings generalize to other scenes. Finally, and perhaps most troublesome of all, the resolution of Yarbus' eye monitor does not permit one to disentangle selection of gaze information from selection of other facial features. Thus, the study by Yarbus is suggestive that people may prefer to look at the eyes of others when several people are depicted in a scene, or, alternatively, when there is activity in the scene, but his study is far from conclusive on this issue.
The aim of the present study was to answer four main questions regarding the selection of gaze information. First, do people prioritize the eyes of people when gaze information is embedded in a number of different complex visual scenes? Second, is the selection of gaze information affected by the task given to observers? Third, does variation in social content of a scene impact the selection of gaze information? Fourth, does variation in the activity level within a scene influence gaze selection?

To get at these questions we presented participants with 20 complex real world scenes that contained either 1 person or 3 persons. The actors in these scenes were either doing something (e.g., reading a book; Active scenes) or were doing nothing (e.g., just sitting on their own; Inactive scenes). Participants were given one of three possible tasks. For one group, participants were asked to simply look at the scenes that they were shown (Look task). As the participants knew that they were being eye monitored (and thus they knew that where they were looking was of interest to the study), we considered this to be the most neutral possible task instruction. Therefore the Look task provided a baseline against which to compare the other task instructions. Participants in a second group were asked to describe the scene (Describe task). Note that, like the Look task, this instruction does not emphasize any particular aspect of the scenes. Participants in a third group were asked to describe where people in the scene were directing their attention (Social Attention task). Thus, the task was again to describe the scene but now the instruction reflected the folk understanding that people look to the eyes in a scene to understand where social attention is being committed (see also Smilek, Birmingham et al. 2006).

Method

Participants

Thirty-nine undergraduate students from the University of British Columbia participated. All had normal or corrected to normal vision, and were naïve to the purpose of the experiment. Each participant received course credit for participation in a one-hour session.
Apparatus

Eye movements were monitored using an Eyelink II tracking system. The on-line saccade detector of the eye tracker was set to detect saccades with an amplitude of at least 0.5°, using an acceleration threshold of 9500°/s² and a velocity threshold of 30°/s.

Stimuli

Full color digital photos were taken of different rooms in the UBC Psychology building. Image size was 36.5 x 27.5 (cm) corresponding to 40.1° x 30.8° at the viewing distance of 50 cm, and image resolution was 800 x 600 pixels. Each of the 20 scenes used contained either 1 or 3 persons, either doing something (e.g., sitting and playing cards; Active scenes) or doing nothing (e.g., just sitting on their own; Inactive scenes). Examples of these scene-types are presented in Figure 3.2a. Due to differences in the number of people (1 or 3) and variation in distance between the people and the camera, the eye region varied in area from 1.69 deg² to 35.38 deg², with an average area of 6.66 deg². Specific experimental details are presented below.

--Figure 3.2--

Procedure

Participants were seated in a brightly lit room, and were placed in a chin rest so that they sat approximately 50 cm from the display computer screen. Participants were told that they would be shown several images, each one appearing for 15 seconds. Each participant was randomly assigned to one of three tasks. The Look group was told to simply “look at” each image. The Describe group was told to “look at, and then describe” each image. The Social Attention group was asked to "describe where the people in the scene are directing their attention". The Describe and Social Attention groups were given an answer booklet, with space available for answering their assigned question for each picture in the order presented. Participants were told that they would have to write their answer for any given picture after the trial was over, i.e., after the image disappeared, and that they could take as long as they needed to write their answer.
Before the experiment, a calibration procedure was conducted. Participants were instructed to fixate a central black dot, and to follow this dot as it appeared randomly at nine different places on the screen. This calibration was then validated with a procedure that calculates the difference between the calibrated gaze position and target position and corrects for this error in future gaze position computations. After successful calibration and validation, the scene trials began.

At the beginning of each trial, a fixation point was displayed in the centre of the computer screen in order to correct for drift in gaze position. Participants were instructed to fixate this point and then press the spacebar to start a trial. One of 20 pictures was then shown in the center of the screen. Each picture was chosen at random and without replacement. The picture remained visible until 15 seconds had passed, after which the picture was replaced with the drift correction screen. During this time participants in the Describe and Social Attention groups wrote an answer using the booklet provided. This process repeated until all pictures had been viewed.

Results

For each image, an outline was drawn around each region of interest (e.g., "eyes") and each region’s pixel coordinates and area were recorded. We defined the following regions in this manner: eyes, heads (excluding eyes), bodies (including arms, torso and legs), foreground objects (e.g., tables, chairs, objects on the table) and background objects (e.g., walls, shelves, items on the walls). Figure 3.2b illustrates these regions. Regions were pooled, such that there was one composite “eye” region made up of all eye regions, one composite “head” region made up of all head regions, etc.

To determine what regions were of most interest to observers we computed fixation proportions by dividing the number of fixations for a region by the total number of fixations over the whole display. These data were area-normalized by dividing the proportion score for each region by its area (Smilek, Birmingham, Cameron, Bischof & Kingstone, 2006; Birmingham et
al., 2008). Note that in doing so we also corrected for changes in the area covered by eyes across scenes with different numbers of people.

To determine where observers’ initial saccades landed in the visual scene, we computed the number of first fixations and second fixations that landed in a region (initial fixations). We also computed the time it took observers to fixate a region for the very first time. To ensure that outliers did not skew the mean effects, we excluded first-fixation latencies that followed display onset by less than 100 ms or more than 2 seconds.

We submitted the fixation proportion data to a 3x2x2x5 mixed analysis of variance (ANOVA) with Task (Look, Describe, Social Attention) as the between-subjects factor and People (1 person vs. 3 people), Activity (Inactive vs. Active) and Region (Eyes, Head, Body, Foreground, Background) as within-subjects factors.

**Question 1: Is gaze information preferentially selected from complex scenes?**

Figure 3.3 shows these data for each region as a function of Task. Looking at this Figure it is evident that for each group eyes were fixated far more often than any other region, as reflected by a main effect of Region, $F(4,144)=542.10, p<0.001$. Pairwise comparisons (Fishers LSD $p<0.05$) revealed that the eyes were fixated the most ($M=0.62$), followed by heads ($M=0.28$), bodies ($M=0.05$), background ($M=0.03$), and finally other objects ($M=0.02$). The initial fixation data (Figure 3.4) showed that this preference for eyes emerged early on, with fixations being more likely to land on the eyes or heads than any other region (main effect of Region, $F(4,144)=34.65, p<0.0001$; Fishers LSD pairwise comparisons, $p<0.05$). This finding dovetails with the latency to first fixate a region, with observers fixating eyes and heads equivalently ($M=499$ ms) and significantly sooner than any other region ($M=923$ ms; main effect of Region, $F(4, 142)=18.46, p<0.0001$; Fishers LSD pairwise comparisons, $p<0.05$).
Question 2: Does the task of describing social attention drive fixations to the eyes?

It is also clear from Figure 3.3 that eyes were fixated more often in the Social Attention task than the Look or Describe tasks, resulting in a Task x Region interaction, F(8,144)=3.73, p<0.0001. Interestingly, while there were greater fixations to the eyes for the Social Attention task (Fishers LSD, p<0.05) than the Describe or Look tasks (which did not differ from each other, p>0.05), fixations to the head regions were similar in all three tasks, whereas there were significantly fewer fixations to the other regions (e.g., body, foreground objects, background) for the Social Attention task relative to the other tasks (Fishers LSD, p<0.05). The initial fixation data (Figure 3.4) showed that the interest in eyes and heads did not differ as a function of task early on (F<1). Thus, the increased preference for eyes in the Social Attention task appeared to be strategic in nature, emerging after an initial interest in eyes that was equal across tasks.

Questions 3 & 4: Does social content or activity affect gaze selection?

Recall that based on Yarbus' Repin study we had hypothesized that increasing the social content of a scene or increasing the activity within a scene, might drive people to look more at the eyes. Figure 3.5 presents the data that address this issue. Here we see that, indeed, increasing the social content of a scene, by adding more people to it, increases fixations to the eyes, but only when there is activity occurring within a scene. This is reflected by a significant People x Activity x Region interaction, F(4,144)=3.17, p<0.05, and confirmed by a one-tailed pairwise comparison (p<0.05). When there is no activity in the scene, there is no effect of adding people (p>0.05). Note that these findings did not vary as a function of group and that there were no higher order interactions in our analyses.

--Figure 3.5--

Discussion

The present study set out to answer four main questions. First, do people prioritize the eyes of people when gaze information is embedded in a number of different complex visual scenes? Second, does the task of describing attentional states in the scene enhance fixations
to the eyes? Third, does variation in the social content of a scene impact the selection of gaze information? Fourth, does variation in the activity in a scene influence gaze selection?

On the first question, the results were clear: observers preferentially selected the eyes over other regions of a scene. This was evident both in terms of fixation proportions and initial fixation data. Fixation proportions revealed that for all task conditions - Look, Describe, and Social Attention - the eyes were fixated most frequently, followed by heads, bodies, background, and finally other objects. Thus, the preferential bias for the eyes of a person persists in real world scenes containing other body parts and objects. That said, it is also clear that eyes did not entirely dominate observers’ attention in these scenes, as fixations did frequent (M=0.38) other body parts and objects in the visual scenes. Finally, the initial fixation data supported observers’ overall preferential bias for gaze information, with participants more likely to fixate the eyes and heads than any other region.

The second question we investigated was whether task would impact the selection of gaze information. Based on the folk understanding that people select gaze information in order to understand the social attention of others, we had predicted that observers would fixate gaze stimuli more often in the Social Attention task than in the baseline Look or Describe tasks. The results supported this prediction. We found a higher fixation proportion for eyes in the Social Attention task than in the Look or Describe tasks. However this task enhancement of gaze selection was not immediate, with the initial fixation being committed most often to the eyes and heads equally across tasks. This suggests that initially observers’ attention was captured by the eyes and heads of people in the scene regardless of task, and that with time the Social Attention group looked more often at the eyes in order to complete their task. It should be noted, however, that we are not claiming that the eyes would be preferentially selected in every task. For instance, it is highly probable that the eyes would be fixated much less if the task were to memorize what the people in the scene were wearing (e.g., Yarbus, 1967). However, a conclusion from the present study is that observers have a natural preference to select the eyes (Look task), and that this preference is later enhanced when they are asked to infer the
attentional states of people in the scene (Social Attention task) but not when they are asked to describe the scene (Describe task).

Finally, we had speculated that Yarbus' (1967) Repin study raised the possibility that increases in the social content of a scene, or the activity of a scene, would lead people to increase their fixations on the eyes in a scene. Our findings supported this prediction when both factors were covaried. That is, when the social content of a scene was increased by adding more people to a scene and when there was activity within a scene, then there was a significant increase in the fixations committed to the eyes. Thus our data confirm the validity of Yarbus' Repin study and resolves a number of long-standing concerns related to that study. Were Yarbus' data an artifact of people sharing knowledge of that particular Repin painting? No, apparently not, as the findings generalize to our very different and varied complex real-world scenes and across all task sets. Were the participants in the Repin study looking at the eyes or at the heads in Yarbus' study? Our data suggest that it was the eyes. And are these fixations on the eyes being driven by the social content or activity in the scene? Our study indicates that it is the interaction between these factors. It is our speculation that this interaction reflects the importance of gaze information for understanding social interactions, that is, actions within social situations. This interpretation is reinforced by our finding that participants fixate the eyes by far the most when the task is to describe the social attention within a scene.

It is worth noting that our present study suggests a subtle, yet potentially very powerful way to examine observers' sensitivity to changes in social content. For instance, it would be instructive to examine how people that are thought to have atypical social attention, such as individuals with autism, scan scenes with one versus many people; and how their exploration of these scenes is affected by the action within it. For instance, if individuals with autism are averse to an increase in social content, as has recently been suggested by Dalton et al. (2005) then the clear prediction is that they will tend to look away from the eyes as people are added to a scene and their activity increases.
Another avenue that seems ripe for investigation is how the factors that impact the selection of gaze information influence the shift of attention to gazed-at locations and objects. A number of possibilities exist. One is that whatever factors drive people to select a gaze stimulus, once gaze is selected, people will tend to shift their attention -- overtly or covertly, or both -- to the gazed-at location. Such an outcome would be in keeping with the current thinking on social attention, which has focused largely on the allocation of attention to a gazed-at location, and has concluded that the shift is largely automatic (Langton et al., 2006; Friesen & Kingstone, 1998). An alternative possibility, however, is that because social content and activity drive people to fixate the eyes, they will actually work against attention being shifted to where gaze is being directed. That is, when gaze stimuli are highly engaging, such as when a scene contains social action, observers will be less likely to disengage from the eyes and shift their attention to gazed-at locations. This outcome would dovetail with recent studies suggesting that many of the lab-based studies of social orienting to gaze direction have overestimated the impact of gaze direction on the orienting of spatial attention (see Kingstone et al., 2003). Specifically, these studies have shown that other directional cues, like arrows, will trigger shifts of attention that closely approximate the effect observed for eyes. The implication is that the eyes and other directional cues will trigger an orienting effect that appears to be automatic when the testing environment is highly impoverished. Whether eyes will trigger a similar effect when the environment is complex is very much an open and important research issue.

We began our report by observing that in examining the effect that gaze stimuli have on the orienting of spatial attention, and the neural systems that may mediate these shifts, past studies of social attention have routinely preselected and isolated the eyes and face stimuli from all other body parts and objects within a scene. We noted that this research approach, while productive in its own right, failed to inform researchers whether gaze was selected preferentially in complex real world social scenes. Nor did this work tell researchers what factors, if any, were critical to this selection process. The present study has taken several important steps on these issues. We have shown unequivocally that while people will fixate other body parts and non-
body objects that are depicted within complex visual scenes, their preferential bias is to fixate
the eyes of others. Our data also suggest that gaze selection is driven by the goal to extract
social attention information, and factors that may change the content of this information, such as
changes in the number of people and activity level within a scene will in turn affect the degree to
which gaze is selected. Our study has also resolved a number of long-standing issues that had
undermined the original classic work of Yarbus. Finally, we have found that our investigation has
broad implications for future investigations and theories of social attention.
Figure 3.1. Scanpaths produced in Yarbus' (1967) study, in which participants freely viewed a set of images. A. Scanpaths for an individual face, showing a selective preference for the eyes. B. Scanpaths for a face accompanied by the rest of the body. Note that the preference for eyes is greatly reduced compared to when viewing an individual face. C. Scanpaths for a social scene, Repin’s *An Unexpected Visitor*. Here a preference for faces/eyes is again observed. From Eye Movements and Vision, by A. L. Yarbus, 1965 (translated by B. Haigh, 1967), New York: Plenum Press. Copyright 1965 by Springer Science and Business Media (pp. 174, 180, and 189).
Figure 3.2. Examples of the four scene types. From top to bottom: 1-person Active, 1-person Inactive, 3-people Active, 3-people Inactive. In Active scenes, the actors were involved in some kind of action (e.g., reading, playing cards, conversing). In Inactive scenes, the actors were sitting quietly on their own. B. Corresponding regions of interest used in analysis (eyes, head, body, foreground objects, background).
Figure 3.3. Fixation proportion data plotted as a function of Task and Region. Observers fixated the eyes the most, followed by heads, bodies, background, and foreground objects. The Social Attention group showed enhanced fixations to the eyes, but decreased fixations to the bodies, foreground objects and background regions relative to the baseline tasks. * indicates a significant difference in means using pairwise comparisons (p<0.05).
Figure 3.4. Proportion of second fixations falling on eyes, heads, bodies, foreground objects, or background, as a function of task (Look, Describe, Social Attention). The eyes and heads were more likely than any other region to receive a second fixation.
Figure 3.5. Fixation proportion data plotted as a function of People, Activity, and Region. Fixations to eyes were enhanced by increasing social content (i.e., 3 people scenes vs. 1 person scenes) when the scenes contained activity (Active scenes). * indicates a significant difference in means using pairwise comparisons (p<0.05).
References


CHAPTER 4

Is there a default bias to look at the eyes?
Human eyes convey a wealth of information that we use on an everyday basis. On a general level, the eyes contain information used to determine someone’s identity, race, age, and gender. At a more social level, the eyes tell us about someone’s attentional, emotional, and mental states (e.g., beliefs, desires). The eyes also facilitate social interactions, by signalling social interest, controlling conversation turn-taking, and exhibiting social dominance or appeasement. For these reasons, the eyes are thought to be extremely informative stimuli (Emery, 2000), conveying a ‘language’ of their own (Baron-Cohen et al., 1997).

The notion that eyes are informative has been bolstered by studies showing that observers can accurately judge someone’s identity and their emotional and mental states based on eye information alone (Baron-Cohen et al., 1997; Langdell, 1978). For example, Baron-Cohen et al. (1997) found that observers were just as accurate at judging complex mental states (e.g., guilt, thoughtfulness, arrogance) from eyes alone as they were at judging these states from a whole face. Similarly, Langdell (1978) found that healthy observers perceived the upper portions of a face (i.e., eyes) to be more informative than the lower regions of the face (i.e., mouth) for recognizing the faces of their peers. Furthermore, eye tracking studies have shown that observers fixate the eyes of a face more often than other facial features while judging emotional expression (Pelphrey et al., 1998), understanding speech (Vatikiotis-Bateson et al., 1998) and recognizing a face (Henderson, Williams & Falk, 2005; Walker Smith, Gale & Findlay, 1977). Thus, there is strong evidence to support the idea that people fixate the eyes relative to other facial features in order to extract information about the identities, emotions, and complex mental states of other people.

What happens when faces are presented along with their other body parts as well as other objects from the environment? For these more complex situations the evidence is remarkably scant, and conflicting. Looking back at the seminal work of Yarbus (1967) one finds that when observers are presented with a full-bodied picture, the bias that one sees toward the eyes of an isolated face (Figure 4.1a) is much less clear-cut (Figure 4.1b). However, it is difficult to know exactly what should be made of this type of observation because the data are
presented in a way that precludes anything more than a rudimentary and qualitative assessment that a particular area appears to be fixated more than another area. What one does not know is how much different regions are fixated (e.g., in terms of fixation frequency or fixation duration), or the order in which different regions are fixated.

To address this fundamental shortcoming, Birmingham et al. (2008b) conducted a series of studies with pictures of one or three people embedded in natural scenes. This work demonstrated that people prefer to look at the eyes of people, e.g., as measured by the proportion of fixations committed to the eyes relative to other regions in the scenes. In addition, Birmingham et al. discovered that this preference for eyes persisted across three different task instructions ("look at the scenes", "describe the scenes" or "describe where the people in the scene are directing their attention"); but that the preference for the eyes was especially pronounced when the task was to report where people in the scene were directing their attention (social attention instruction). These results suggest that there exists a fundamental "default" bias toward looking at the eyes of people in a scene that can be enhanced by the intention to extract information about the social states of others. In other words, while task instruction can modulate people's tendency to look at the eyes of others, all things being equal, there is a basic tendency to look at the eyes of others. This interpretation of the data is highly consistent with models of social attention which propose that humans have a special sensitivity to eyes, and eye-like stimuli, and that we have evolved to preferentially select eyes from our environment (e.g., Baron-Cohen, 1994).

Looking at the study by Birmingham et al. (2008b) with a critical eye, however, one could challenge this interpretation of the data and argue that the data merely show that people will look at what you tell them to look at. In other words, participants in the study may have merely interpreted the "look", "describe" instructions to mean "look at the people", and "describe the people" (and indeed, it is clear that the "social attention" instruction explicitly demanded that observers look at the people) in the scenes.
Support for this alternative explanation, which rejects the idea of a "default preference for the eyes of others", can be found in the seminal work of Yarbus (1967). He found that when participants were presented with Repin's painting, "An unexpected visitor" (Figure 4.1c), the scene regions that were fixated depended strongly on the task that was assigned. When asked to estimate the wealth of people in the scene, fixations were directed mostly at the clothing and objects in the scene; when asked to remember the people's clothes, fixations were committed to their clothes; when asked to remember the positions of people and objects in the room, fixations were distributed widely over the scene; and when asked about the people's ages, or to estimate how long the unexpected visitor had been away from the family, fixations were directed mostly at the faces. Thus, the work of Yarbus could be interpreted as suggesting that there is no fundamental bias to look at eyes of people in scenes; rather people look at the eyes when eyes are relevant to the experimental condition to which they are assigned. In other words, scene scanning depends on the task.

Yet, as with Yarbus' other work, there are several important limitations to this study that demand caution when trying to draw firm conclusions. One is that in all of Yarbus' task conditions observers did look at the heads/eyes. Exactly where they looked on the head (e.g., eyes vs. elsewhere), how much, and when these fixations occurred however is unknown; for while Yarbus' scanning sessions were very long (3 minutes), the actual location, frequency, time and order that specific fixations occurred is not reported. It might be the case, for example, that even in the scenes when there were relatively few fixations on the heads/eyes, these fixations occurred early in the scanning session. This would be consistent with a default bias toward looking at the face and eyes. Thus, what one needs are measures of overall fixation preferences over the whole course of viewing.

There are additional concerns as well, perhaps the most noteworthy one being that there was only one person tested across the different task conditions. In other words, the extent to which fixation performance in the different tasks reflects effects that are general across individuals rather than specific to the particular individual tested, is unknown.
In sum, while Yarbus’ seminal data are suggestive that a preference for looking at the eyes of faces in a scene might simply be a reflection of the task, there are sound reasons to question this idea as well.

To resolve this issue, the present study presented different groups of participants with different task instructions. Three groups were asked questions that were unambiguously about the people in the scenes: participants were asked to either study the attentional states of people in the scene (e.g., where their attention was committed), the emotional states of people in the scene (e.g., what they were feeling), or the cognitive states of people in the scene (e.g., what they were thinking). Two other groups were given task instructions that were intentionally stated to be general and to avoid explicitly drawing attention to the people in the scenes: one group was asked to study the meaning of the scenes, and the other group was instructed to study the informative aspects of the scenes. Nevertheless, it may be that observers normally perceive people to be inherently meaningful and informative to scenes, so, a final group was instructed to study the uninformative aspects of the scenes.

If observers preferentially select the eyes of people in scenes only when they are instructed to study the people, then we predicted that relative to the other regions in the scenes, there will be a clear bias to fixate the eyes for the first three “social” groups, that this bias will be absent for the last "uninformative" group, and there would be a preferential bias for the eyes that fell between these two extremes for the "meaning" and "informative" groups (the key for these two groups of course may be the extent to which they understand people to be meaningful and informative).

On the other hand, if there is a default bias for the eyes that is fundamental across observers, it should be relatively insensitive to task instructions, with participants in all groups demonstrating a preferential bias for the eyes relative to all other regions in the scene. This does not mean, however, that instruction should have no effect. Based on Birmingham et al (2008) one expects task instructions that explicitly instruct participants to consider the people in the scenes should enhance the bias toward to eyes. But again, all groups, even the participants
in the "uninformative group" should now demonstrate a fundamental bias toward the eyes relative to all other regions.

We computed overall fixation preferences (fixation proportions) and the location of the first fixation, to determine not only which regions were preferred over time but also which ones captured attention initially.

Method

Participants

Fifty-five undergraduate students from the University of British Columbia participated (42 female, 13 male). All had normal or corrected to normal vision, and were naïve to the purpose of the experiment. Each participant received course credit for participation in a one-hour session.

Apparatus

Eye movements were monitored using an Eyelink II tracking system. The on-line saccade detector of the eye tracker was set to detect saccades with an amplitude of at least 0.5°, using an acceleration threshold of 9500°/s² and a velocity threshold of 30°/s.

Stimuli

Full color digital photos were taken of different rooms in the UBC Psychology building. Image size was 36.5 x 27.5 (cm) corresponding to 40.1° x 30.8° at the viewing distance of 50 cm, and image resolution was 800 x 600 pixels.

Twelve scenes (3 rooms, 4 scene types) were presented. These scenes contained a variety of social situations containing 1 or 3 persons. All scenes were comparable in terms of their basic layout: each room had a table, chairs, objects, and background items. Examples of these scene-types are presented in Figure 4.2.

Procedure

Participants were seated in a brightly lit room, and were placed in a chin rest so that they sat approximately 50 cm from the display computer screen. Participants were told that they would be shown several images, each one appearing for 10 seconds.
All participants were told that there would be a later memory test. That is, all participants were given the following instruction: “In this experiment you will be shown a set of photographs of scenes. I want you to study these scenes for a later memory test.” However, no memory test was actually administered, as we were interested in how viewing patterns would differ according to the instructions given for studying the scenes. Each participant was randomly assigned to one of six instructions: ‘study the attentional states of people in the scene’; ‘study the emotional states of people in the scene’; ‘study the cognitive states of people in the scene’; ‘study the meaning of the scene’; ‘study the informative parts of the scene’; or ‘study the uninformative parts of the scene’. After the study session, participants were debriefed and given a questionnaire asking them about their impressions of the experiment.

Before the experiment, a calibration procedure was conducted. Participants were instructed to fixate a central black dot, and to follow this dot as it appeared randomly at nine different places on the screen. This calibration was then validated with a procedure that calculates the difference between the calibrated gaze position and target position and corrects for this error in future gaze position computations. After successful calibration and validation, the scene trials began.

At the beginning of each trial, a fixation point was displayed in the centre of the screen in order to correct for drift in gaze position. Participants were instructed to fixate this point and then press the spacebar to start a trial. A picture was then shown in the centre of the screen. Each picture was chosen at random and without replacement. The picture remained visible until 10 seconds had passed, after which the picture was replaced with the drift correction screen. This process repeated until all pictures had been viewed.

Results

For each image, an outline was drawn around each region of interest (e.g., "eyes") and each region’s coordinates and area were recorded. We defined the following regions in this manner: eyes, heads (excluding eyes), bodies (including arms, torso and legs), foreground
objects (e.g., tables, chairs, objects on the table) and background objects (e.g., walls, shelves, items on the walls). Figure 4.2 illustrates these regions.

**Fixation proportions**

To determine what regions were of most interest to observers we computed fixation proportions by dividing the number of fixations for a region by the total number of fixations over the whole display. These data were area-normalized by dividing the proportion score for each region by its area (Birmingham et al. 2008a,b).

Figure 4.3 shows the fixation proportions for the scenes as a function of Instruction (Attention, Emotion, Cognitive, Meaning, Informative, Uninformative) and Region (eyes, heads, bodies, foreground objects, background).

Recall that the critical prediction was that if there is a default bias to select the eyes, then all groups, including the uninformative group, should look preferentially at the eyes relative to the other regions. Figure 4.3 clearly shows that this was the case. In addition, as predicted, it is clear that the bias to fixate the eyes was enhanced by the social tasks (attention, emotion, cognitive).

To confirm these impressions, the data were submitted to a mixed ANOVA with Instruction (Attention, Emotion, Cognitive, Meaning, Informative, Uninformative) as a between-subjects factor and Region (eyes, heads, bodies, foreground objects, background) as a within-subjects factor. This analysis revealed a highly significant effect of Region (F(4,196)=673.51, p<0.0001), reflecting that overall participants preferred to scan the eyes more than any other region (Fishers LSD, p<0.05). Separate tests on the individual task groups confirmed that each group showed a general preference to look at the eyes (Fishers LSD, p<0.05). However, an Instruction x Region interaction (F(20,196)=2.30, p<0.05) reflected that the eyes were fixated more frequently by the Attention, Emotion, and Cognitive groups than by the Meaning, Informative, and Uninformative groups (Fishers LSD, p<0.05).
First Fixation

We were interested in determining how early the general bias to look at eyes emerges. To this end we examined the first fixations made by viewers to a scene (i.e., the first fixation after the experimenter-determined fixation at centre).

We computed the proportion of first fixations that landed in each region, as a function of Instruction (Figure 4.4). A mixed Instruction (Attention, Emotion, Cognitive, Meaning, Informative, Uninformative) by Region ANOVA revealed a main effect of Region ($F(4,196)=64.56, p<0.0001$), reflecting that eyes and heads were both most likely to be fixated first than any other region (eyes: 0.35; head: 0.34; background: 0.19; foreground objects: 0.07, bodies: 0.04) (Tukey Kramer, $p<0.05$). The background was the next most likely to be fixated relative to the bodies and foreground objects ($p<0.05$). However, there was no interaction between Instruction and Region ($F(20,196)=1.20; p>0.10$), indicating that the groups did not differ in where they directed their first fixation.

---Figure 4.4--

Cumulative fixation proportions

The results from the initial fixation data suggest that there is a default bias to look at the eyes that exists equally across tasks within the first fixation. Thus, any effects of strategy on viewing preferences must have manifested later on in the viewing period. To confirm this possibility, we computed cumulative fixation proportions for the eyes in 1-second time steps for the 10 seconds of display duration (Figure 4.5). These data were area-normalized by dividing the proportion score by the area of the eye region.

---Figure 4.5--

As can be seen from Figure 4.5, the task groups looked equally often at the eyes until about 8 seconds, at which time the three social groups started to diverge from the three other groups. To confirm these impressions we conducted a mixed ANOVA with Instruction as the between-subjects factor and Interval as the within-subjects factor. This analysis revealed a main effect of Interval ($F(9,441)=849.12; p<0.0001$), reflecting that the eyes were re-fixated over
time, resulting in an overall increase in the cumulative fixation proportions. In addition, there was an Group x Interval interaction (F(45, 441)=5.12; p<0.001), reflecting that the task groups diverged over time in terms of their interest in the eyes. Our analysis revealed that group differences did not begin to emerge until after 8 seconds, at which point there was a significant difference between each of the social groups (attention, emotion, cognitive) and the uninformative group (Tukey-Kramer p<0.05), and between the emotion group and the informative and meaning groups. At 9 seconds each of the social groups was fixating the eyes significantly higher than the three other groups (Tukey-Kramer, p<0.05).

Discussion

The present study asked the following question: Is there a default bias to select the eyes of people in complex social scenes, or is this behaviour an artefact of task instructions that encourage observers to look at people? To test this question, we presented a variety of scenes and task instructions. Three 'social' task groups were asked questions that were clearly about the people in the scenes (study the Attentional, Emotional, or Cognitive states in the scene); two 'general' task groups were asked questions that did not explicitly draw attention to the people in the scenes (study the Meaning of the scene; study the Informative parts of the scene); and one group was asked to study the Uninformative parts of the scene.

If observers preferentially select at the eyes of people in scenes only when they are instructed to study the people, then the prediction is that fixations to the eyes would be strong for the first three groups, absent for the last "uninformative" group, and in the middle for the "meaning" and "informative" groups (assuming these two groups understand people to be meaningful and informative within complex scenes). On the other hand, if there is a default bias for the eyes that is fundamental, it should be relatively insensitive to task instructions, and observers in all groups should select eyes more frequently than any other region. That is, even observers asked to study uninformative aspects of the scene should preferentially look at the eyes of the scene. However, we also strongly expected that the social tasks would enhance this default bias to select the eyes relative to the other tasks, as would be predicted by the findings
of Birmingham et al. (2008). The critical point, however, is that a fundamental bias to select eyes should be reflected to some extent in all tasks.

The results were clear. We found that regardless of task, observers showed a general preference to look at the eyes in the scene relative to other scene regions. In addition, as predicted, overall fixation proportions for the eyes were enhanced by the social tasks (Attention, Emotion, Cognitive) relative to the other tasks (Meaning, Informative, Uninformative). However, all task groups were equally likely to fixate the eyes on the first fixation. Indeed, the eyes and the heads were both most likely to receive the first fixation, suggesting that there is a default bias to look at the eyes (and heads) very early on in viewing. It is only much later on in viewing that task instructions affect the selection of eyes: the social task groups increased their fixations to the eyes relative to the other groups by about 8 seconds (the last 2 seconds of viewing).

A default bias to select the eyes is consistent with models of social attention that purport the eyes to be highly important social stimuli (e.g., Baron-Cohen, 1994). According to these models, the eyes are the primary cue to the attention, intentions, and complex mental states of other people, and as such we are highly sensitive to their presence. Indeed, this sensitivity appears at an early age (e.g., Haith, Bergman & Moore, 1977), and may be supported by specific neural mechanisms (Baron-Cohen, 1994). For instance, infants have been found to preferentially scan a face’s eyes by 2 months of age (Haith, Bergman & Moore, 1977; Maurer & Salapatek, 1976). Before this age, infants spend most of their time looking away from the face (Maurer & Salapatek, 1976) or, when looking at the face, at the periphery of the face (e.g., chin and hairline, Maurer & Salapatek, 1976; Haith, Bergman & Moore, 1977). Baron-Cohen (1994) suggests that this preference for the eyes is supported by a specific module, called the Eye Detection Director (EDD), which both detects the presence of eyes and computes the direction of gaze (this latter ability develops by about 9 months). Detecting eyes and understanding the social information that eyes portray is so important that it is thought that any disruption of the development of this ability will lead to severe social deficits, such as autism spectrum disorders (Baron-Cohen, 1994). Indeed, there is accruing empirical evidence that individuals with autism
fail to orient to social stimuli (Dawson et al., 1998), particularly the eyes of others (Dalton et al., 2005; Klin et al., 2002), and have difficulty inferring the mental states of others from the eyes of a person (Baron-Cohen et al., 1997). Thus, a default bias to select the eyes may be a key element of normal social development.

One possibility is that the default bias to look at eyes might be difficult to overcome without specific instructions to avoid looking at the eyes. Indeed, perhaps a useful future study would be to replicate the task instructions of Yarbus’ (1967) study (e.g., remember the clothing of people in the scene), and to determine whether a default bias to select the eyes still exists or can be overridden. Under such instructions, one might find that the default bias to look at the eyes is expressed, but only in the first few fixations (e.g., an ‘automatic’ tendency to look at the eyes). After the first few fixations it may be that top-down strategies can override the tendency to re-fixate the eyes. Note, however, that in the present study we chose not to instruct observers to look at specific aspects of the scene (e.g., clothing, objects), or to specifically ‘avoid’ looking at the eyes, because we did not want to create unnatural viewing patterns. Instead, we chose to keep the instructions broad, but to vary how ‘people-relevant’ the instructions were. We felt that this would promote less forced viewing patterns and would reveal any pre-existing bias to select at the eyes across tasks.

One implication of the presence of a default bias to select the eyes is that it will normally be present whenever someone views a scene with people. Thus, finding a general preference for eyes for a given task instruction may simply reflect the default bias, and not a strategic selection of eyes for the task at hand. Thus, an important methodological consideration is that the default bias must always be measured in order to infer whether task instructions lead observers to select the eyes above and beyond what would be expected by default. This point is particularly relevant given the plethora of scene perception studies that administer specific task instructions without including a control task (such as a free viewing instruction) (e.g., De Graef et al., 1990; Henderson et al., 1999; Loftus & Mackworth, 1978). In these studies, it is difficult to know whether fixation preferences were a product of the task or whether they were due simply
to fundamental preferences that would be present in all tasks. For instance, a common instruction given to participants in scene viewing studies is to “study the scenes for a later memory test” (Henderson et al., 1999; Loftus & Mackworth, 1978). Many of these studies have found that observers fixate semantically informative scene items more often than semantically uninformative scene items. However, because no free-viewing control instruction was included, one cannot know whether this pattern was due to the memory instruction itself or because of a default preference to look at informative scene items.

In conclusion, the present study has demonstrated that there is a default preference to look at the eyes across a variety of task instructions. Our working hypothesis is that this reflects observers’ understanding that eyes convey important social-communicative information, particularly about complex mental states. Indeed, this idea is supported by the finding that the default bias to look at eyes was enhanced by specific instructions to study the mental states of people in the scenes (Attention, Cognitive, Emotion). However, the finding that all observers showed a general preference to look at eyes suggests that this basic interest in the mental states of others may be fundamentally present. Indeed, it is our belief that inferring the mental states of people in the scene was critical to building an in-depth understanding of the scene itself.
Figure 4.1

Figure 4.1. Scanpaths of an individual face (A), a face accompanied by the rest of the body (B), and a social scene (C). From Eye Movements and Vision, by A. L. Yarbus, 1965 (translated by B. Haigh, 1967), New York: Plenum Press. Copyright 1965 by Springer Science and Business Media (pp. 174, 180, and 189).
Figure 4.2. Examples of the scenes. B. Corresponding regions of interest used in analysis (eyes, head, body, foreground objects, background objects).
Figure 4.3. Fixation proportion data plotted as a function of Region and Instruction. Fixations to eyes were enhanced in the three ‘social’ (Attention, Emotion, Cognitive) groups relative to the three other groups.
Figure 4.4. Proportion of first fixations falling on eyes, heads, bodies, foreground objects, or background, as a function of Instruction (Attention, Emotion, Cognitive). The eyes and heads were more likely than any other region to receive a first fixation. There were no group differences.
Figure 4.5. Cumulative fixation proportions eyes as a function of viewing interval (1-s bins) and Instruction.
References


CHAPTER 5

Remembering social scenes

A version of this chapter will be submitted for publication. Birmingham, E., Bischof, W.F., & Kingstone, A. Remembering social scenes.
As social beings, humans are highly sensitive to the presence of other people’s eyes (Baron-Cohen, 1994; Birmingham et al., 2008(a),(b); Emery 2000). Indeed, because the eyes are an optimal stimulus for communicating the direction of attention (Langton et al., 2000), prominent models of social attention have proposed that detecting eyes and where those eyes are looking is a critical component of normal social development (Baron-Cohen, 1994). For instance, inferring where someone is attending in the environment may be an important precursor for the development of a theory of mind, which includes complex mental state attributions about other people (Baron-Cohen, 1994). As such, it is reasonable to propose that humans have developed a fundamental tendency to preferentially select eyes from the environment (e.g., Baron-Cohen, 1995).

Indeed, such a ‘default bias’ to select the eyes has been demonstrated empirically. In a recent study, Birmingham et al. (under review) showed that when observers were asked to study pictures of natural visual scenes, they looked preferentially at the eyes of the people in scenes regardless of what they were told to study, i.e., the mental states of the people in the scenes, or the informative - or even uninformative – aspects of the scenes. This seemingly default bias to look toward the eyes was accentuated after nearly 10 seconds of scene viewing, with those participants that were told to study the mental states of the people in the scenes committing even more fixations to the eyes than other participants. We have proposed that this default interest in eyes reflects observers’ fundamental interest in the mental states of others (Birmingham et al., under review).

To date, all the studies by Birmingham and colleagues have been concerned with the selection of eyes when a scene is first viewed, that is, when it is being encoded (Birmingham et al., 2000 (a),(b), under review). The present study examined whether this preferential selection of eyes extends to a very different type of situation, that is, one that does not only involve encoding scene information but also retrieving scene information for a memory test. Simply put, while we know that people look at the eyes when they encode social scenes, we do not know that they will look at the eyes when they try to recognize social scenes. This is important for at
least two fundamental reasons. The first is that it is important to determine the extent that the findings and interpretations of the past scene viewing data (e.g., Smilek et al., 2006; Birmingham et al., 2008 (a),(b), under review) generalize to different contexts; for doing so one can begin to assess the robustness and external validity of the data and interpretations that are placed on them. Second, this study will shed light on previous research concerning the utility of eye movements for memory, as different studies make very different predictions as to how participants will perform in our investigation.

One body of research suggests that there is a functional link between eye movements and scene memory (Althoff & Cohen, 1999; Henderson, Williams & Falk, 2005; Hollingworth & Henderson, 2002; Loftus, 1972; Nelson & Loftus, 1980). These studies have found that memory for previously fixated scene items is better than for items that were not fixated (Hollingworth & Henderson, 2002), that changes to memorized scenes are more likely to be noticed when the change is fixated (Henderson, Williams, Castelhano & Falk, 2003; Nelson & Loftus, 1980), that memory for images is better when they are scanned at encoding than when fixations at encoding are restricted (Henderson et al., 2005), and that familiar images are scanned in ways that are distinct from novel images, with eye movements for familiar images restricted to a smaller number of informative scene regions that were fixated at encoding (Althoff & Cohen, 1999; Henderson et al., 2005; Smith et al., 2006). Here, the prediction is that eye movements are important for memory, and that observers will find the same regions to be useful for scene recognition that they find useful for scene encoding. As a result, they will re-fixate the same regions they focused on while encoding the scenes (Althoff & Cohen, 1999; Henderson, et al., 2003). An alternative possibility is that scene recognition can be done quickly and efficiently, that is, without many eye movements (Biederman, 1981; 1988; Biederman, Mezzanotte, & Rabinowitz, 1982; Potter, 1976). Thus while observers in our study may look at the eyes during scene encoding, they may not find this information useful or necessary for scene recognition. As a result observers might demonstrate a marked reduction in fixations to the eyes during
scene recognition relative to scene encoding. The present study aimed to distinguish between these two outcomes.

The design of the study was simple and straightforward. Observers were assigned randomly to two groups. One group was told that they would later be asked to recognize the scenes in a test session (Told group); another group was not informed of the later memory test and simply asked to freely view the images (Not Told group). Both groups were subsequently given a memory test, in which scenes from the pretest session were presented along with scenes that were never seen before.

For the pretest session, based on previous work showing a default bias to look at the eyes upon initially viewing a scene (Birmingham et al., under review), we predicted that observers in both groups would generally look more at the eyes of the people in the scenes than other regions. While we believe that this default bias for eyes reflects observers’ general interest in the mental states of others (Birmingham et al., under review) based on recent pilot data using the current design (Birmingham et al., 2007) we expected observers in the Told group to look at the eyes more than the Not Told group. The key question that we sought to answer with the present study, however, was whether this difference would persist during recognition or be significantly reduced.

Method

Participants

Eighteen undergraduate students from the University of British Columbia participated. All had normal or corrected to normal vision, and were naïve to the purpose of the experiment. Each participant received course credit for participation in a one-hour session.

Apparatus

Eye movements were monitored using an Eyelink II tracking system. The on-line saccade detector of the eye tracker was set to detect saccades with an amplitude of at least 0.5°, using an acceleration threshold of 9500°/s² and a velocity threshold of 30°/s.
Stimuli

Full color digital photos were taken of different rooms in the UBC Psychology building. Image size was 36.5 x 27.5 (cm) corresponding to 40.1° x 30.8° at the viewing distance of 50 cm, and image resolution was 800 x 600 pixels.

Pretest session (15 scenes: 3 rooms, 5 scene types). Twelve of the fifteen study scenes were “People scenes”. These scenes contained a variety of social situations containing 1 or 3 persons. All scenes were comparable in terms of their basic layout: each room had a table, chairs, objects, and background items. Three of the fifteen pretest scenes were “No people scenes”, containing a single object resting on the table. Examples of these scene-types are presented in Figure 5.1.

Test session (56 scenes: 8 rooms, 7 scene types). Thirty-two of the test scenes were “People scenes” as above (12 old, 20 new). Sixteen of the test scenes were “No people scenes”, containing one or three objects resting on the table (3 old, 13 new). Eight additional (new) scenes contained one person doing something unusual, such as sitting with a Frisbee on his head. These scenes were included to keep the participants interested, but were not included in the analysis.

---Figure 5.1---

Procedure

Participants were seated in a brightly lit room, and were placed in a chin rest so that they sat approximately 50 cm from the display computer screen. Participants were told that they would be shown several images, each one appearing for 10 seconds.

Pretest session: Each participant was randomly assigned to one of two instruction groups. The Told group was told that they would be shown 15 images, and that they would be asked to recognize each image in a later memory test. The Not Told group was told to simply “look at” each image, and was not informed of the later memory test. After the pretest session, a brief questionnaire was given to participants asking them about their impressions of the experiment.
Test session: Both groups (Told, Not Told) were informed that they would be shown 56 images, and that they were to view each one and then decide if the image was OLD (i.e., they had seen it in the Study session), or NEW (i.e., they had never seen it before). After an image was presented (for 10 s), a response screen appeared asking them to respond with ‘1’ on the keyboard if they thought the image was OLD, and ‘2’ on the keyboard if they thought the image was NEW. Participants had an unlimited amount of time to respond.

Before the experiment, a calibration procedure was conducted. Participants were instructed to fixate a central black dot, and to follow this dot as it appeared randomly at nine different places on the screen. This calibration was then validated with a procedure that calculates the difference between the calibrated gaze position and target position and corrects for this error in future gaze position computations. After successful calibration and validation, the scene trials began.

At the beginning of each trial, a fixation point was displayed in the centre of the screen in order to correct for drift in gaze position. Participants were instructed to fixate this point and then press the spacebar to start a trial. A picture was then shown in the center of the screen. Each picture was chosen at random and without replacement. The picture remained visible until 10 seconds had passed, after which the picture was replaced with the response screen. After the participant entered a response, the drift correction screen appeared in preparation for the next trial. This process repeated until all pictures had been viewed.

Results

Eye movement data handling

For each image, an outline was drawn around each region of interest (e.g., "eyes") and each region’s coordinates and area were recorded. We defined the following regions in this manner: eyes, heads (excluding eyes), bodies (including arms, torso and legs), foreground objects (e.g., tables, chairs, objects on the table) and background objects (e.g., walls, shelves, items on the walls). Figure 5.1 illustrates these regions.
Fixation Proportions.

To determine what regions were of most interest to observers we computed fixation proportions by dividing the number of fixations for a region by the total number of fixations over the whole display. These data were area-normalized by dividing the proportion score for each region by its area (Birmingham et al., 2008a, b, under review).

Figures 5.2 and 5.3 show the fixation proportions for People (Figure 5.2) and No People (Figure 5.3) scenes, as a function of Instruction (Told, Not Told), Session (Study, Test), and Region (eyes, heads, bodies, foreground objects, background).

---Figure 5.2---

**People scenes**: The data for the People scenes were submitted to a mixed ANOVA with Instruction (Told, Not Told) as a between-subjects factor and Session (Pretest, Test) and Region (eyes, heads, bodies, foreground objects, background) as within-subjects factors. This analysis revealed a highly significant effect of Region (F(4,64)=9.45, p<0.0001), reflecting that overall participants preferred to scan the eyes over any other region. An Instruction x Region interaction (F(4,64)=3.18, p<0.02) indicated that while both groups fixated the eyes more than any other region, the eyes were fixated more frequently by the Told group than the Not Told group (Fishers LSD p<0.05), and that the heads were fixated more frequently by the Not Told group than by the Told group (Fishers LSD p<0.05). Fixation proportions for the other regions did not differ between the two groups. In short, participants are especially likely to look at the eyes when they are simply asked to encode and remember scenes with people. There were no other interactions, (all ps>0.05), including the Instruction x Session x Region interaction (F(4,64)=2.14, p>0.05), indicating that the viewing patterns within each Instruction group did not change from pretest to test sessions. That is, observers in each group fixated the same regions during scene recognition that they fixated during scene encoding.

---Figure 5.3---

**No People scenes**: The data for the No People scenes were submitted to an Instruction (Told, Not Told) x Session (Pretest, Test) x Region (foreground objects, background) mixed
ANOVA. This analysis revealed a main effect of Region (F(1,16)=18.24, p<0.001, with participants preferring overall to scan the foreground objects in scenes without people (Tukey-Kramer, p<0.05). However, unlike with the People scenes, Instruction had no significant effect on scanning patterns, with Instruction x Region returning a nonsignificant result (F(1,16)=2.70, p>0.05). There was also a Session x Region interaction (F(1,16)=51.61, p<0.0001) reflecting that background was fixated more frequently in the pretest session and foreground objects were fixated more frequently in the test session (Tukey-Kramer, p<0.05).

First Fixation

We were interested in determining how early the enhancement of attention to eyes emerged for the Told group relative to the Not told group. Thus, we examined the first fixation made by viewers to the People scenes (i.e., the first fixation after the experimenter-determined fixation at centre).

We computed the proportion of first fixations that landed in each region, as a function of Instruction and Session (Figure 5.4). A mixed Instruction (Told, Not Told) by Session (Pretest, Test) by Region ANOVA revealed a main effect of Region (F(4,64)=29.45, p<0.0001), reflecting that heads were more likely to be fixated first than any other region (head: 0.43; eyes: 0.29; background: 0.16; bodies: 0.08; foreground objects: 0.04) (Tukey Kramer, p<0.05). Eyes were the next most likely to be fixated first, significantly more so than background, foreground objects, and bodies. There was also a significant Instruction x Session x Region interaction (F(4,64)=3.07, p<0.05). As can be seen from Figure 5.4, this higher order interaction indicated that the Told group was more likely to fixate the eyes first than the Not Told group, both in the pretest and test sessions (Tukey-Kramer, p<0.05). In contrast, the Told group was less likely to fixate the heads than the Not Told group, but only in the study session (Tukey-Kramer, p<0.05).

Recognition Accuracy

Although we were not explicitly interested in recognition performance, we computed recognition accuracy to confirm that participants were engaged in the task and were able to
recognize the scenes. They were. There were no group differences in accuracy on the test session, with both groups performing very well (Told group mean accuracy: 95.3%; Not Told group mean accuracy: 95.4%). These high accuracy scores were expected given the evidence that people are excellent at recognizing even very large numbers of scenes (e.g., Standing, 1973).

Discussion

Two groups were tested in the present study: one was told that they would later be asked to recognize the scenes in a test session (Told group); the other was not informed of the later memory test and simply asked to freely view the images (Not Told group). Both groups were subsequently given a memory test, in which scenes from the pretest session were presented along with scenes that were never seen before. We predicted that observers in both groups would generally look more at the eyes of the people in the scenes than other regions (Birmingham et al., under review) but that observers in the Told group would look at the more eyes than the Not Told group (Birmingham et al., 2007). The key issue was whether this difference would persist during recognition or be significantly reduced.

The results were clear. For the No People scenes the Told group and the Not Told group had very similar fixation patterns: overall, both groups preferentially fixated the foreground objects, especially during scene recognition. There were no group differences.

For the people scenes a very different pattern emerged. Both groups preferentially selected the eyes relative to other regions in the scenes. Moreover, fixation proportions revealed that observers who were asked to study the scenes (Told group) fixated the eyes more frequently than observers who were asked to freely view the scenes (Not Told group). The implication is that when simply asked to study the People scenes, observers select the eyes to extract information that they perceive as useful for remembering the scenes. As we will discuss, we believe that observers may have found mental state information to be particularly useful for recognizing the scenes. Interestingly, the difference between Told and Not Told groups in the present study re-emerged in the test session. The Told group fixated the eyes just as strongly in
the test session as in the pretest session, and again more frequently than the Not Told group in the test session. Similarly, the Not Told group’s fixation preferences did not differ significantly between pretest and test. This suggests that the information selected from the scenes during initial scanning, both when studying and simply looking at the scenes, has a fundamental behavioural significance in that it will be applied by participants when they are asked to remember what they have previously seen.

Interestingly, the effect of instruction on viewing patterns was evident within the first fixation. That is, the Told group was more likely to fixate the eyes on the first fixation than was the Not Told group, both at pretest and at test. The early effects of instruction are particularly interesting because Birmingham et al. (under review) did not find that early fixations were affected by task. That is, in the study by Birmingham et al. (under review) the groups who studied the meaning of the scene, the informative aspects of the scene, and the uninformative aspects of the scene, were just as likely to fixate the eyes on the first fixation as were the three groups who studied the mental states of the people in the scenes. It was only later on that the three social/mental state study groups enhanced their fixations to the eyes relative to the other groups. It was based on those results that the authors suggested that the default bias to select the eyes is initially equal across tasks, and that task enhances this bias much later on in viewing. The fact that a difference in first-fixations to the eyes emerged between the Told and Not Told groups in the present study, but not between the groups in Birmingham et al.’s (under review) investigation, most likely reflects the fact that in Birmingham et al.’s study all groups were told to study the scenes for a later memory test. In other words, when participants have the intention of studying the scenes, there is an enhanced first-fixation bias to the eyes regardless of what specific aspect of the scene they are asked to study.

In sum, the present study has shown that the default bias to select the eyes can be accentuated by the instruction to study and remember the scenes. Why might this be? One possibility is that observers perceive the mental states of people in the scene to be particularly useful for constructing an accurate and meaningful scene representation (Birmingham et al.,
under review). This is supported by the finding from previous studies that observers look more at the eyes when the scene is high in social content and activity, suggesting that observers become more interested in attentional states of people in the scene in order to understand the central social activity in a scene (Birmingham et al., 2008(a), (b)). Indeed, we have already discussed our finding that observers look more to the eyes when asked about the mental states of people in a scene, suggesting that the eyes are perceived to be informative in this regard (Birmingham et al., 2008b, under review). An alternative possibility is that observers chose to memorize the identities of the people in the scene as their strategy. However, we think this is unlikely for several reasons. First, the identities of people were repeated across scenes, and so simply remembering ‘who’ was in each scene would not be sufficient for recognizing the scene.

In addition, if observers in the Told group were trying to memorize the identities of the people in the scene, we would have expected fixations on both eyes and heads to be greater than the Not Told group. In contrast, enhancement for the Told group was specific to the eyes, with the Not Told group fixating the heads more so than the Told group. Thus, we do not believe observers were simply memorizing the identities of people in the scene. Rather, we suggest that they were extracting social information from the eyes, such as the attentional and cognitive states of the people in the scene, to construct meaningful scene representations.

Moreover, we have found that the selection biases that are expressed during an initial scene viewing of different visual scenes will re-express themselves during a subsequent memory task. This has several implications. First, it makes the interesting point that it is not simply the case that any task will enhance selection of the eyes – i.e., a surprise memory test did not lead the Not Told group to fixate the eyes more frequently than they did in the pretest session. In addition, the findings are consistent with the literature showing that eye movements are functional in scene memory (Althoff & Cohen, 1999; Henderson, Williams & Falk, 2005; Hollingworth & Henderson, 2002; Loftus, 1972; Nelson & Loftus, 1980). Indeed, the findings support that when remembering a scene, observers tend to re-fixate the regions that they selected when learning a scene (Althoff & Cohen, 1999; Henderson et al., 2005). While the
present data do not speak to the issue of whether observers could have successfully recognized the scenes without re-fixating the regions selected during encoding (e.g., Biederman, 1981; 1988; Biederman, Mezzanotte, & Rabinowitz, 1982, Potter, 1976), the results certainly suggest that observers who are remembering a scene show the tendency to repeat their viewing preferences from the time of encoding.

Finally, an intriguing finding for future investigation is that, unlike the No People scenes, for the People scenes observers rarely fixated the foreground objects, relative to the eyes and heads. We have argued here that this supports the idea that typically developing humans have a special sensitivity to people and their eyes as sources of social communicative information (Baron-Cohen, 1994). But how would atypically developing individuals, such as individuals with autism spectrum disorder (ASD), perform when asked to perform the same task as the Told group? Given that individuals with ASD have been found to have an aversion to social stimuli, particularly to people and their eyes (e.g., Dalton et al., 2005; Pelphrey et al., 2002), the prediction is that individuals with ASD will rely more on the foreground objects when encoding and retrieving social scenes from memory, and less on the eyes. Comparison of this fixation pattern relative to their viewing preferences for No People scenes would shed light on what kinds of information individuals with ASD naturally prioritize in real world social scenes, and what role aversion to social stimuli plays in their preference for nonsocial stimuli.
Figure 5.1. (A) Examples of People scenes and No People scenes used in the experiment. (B) Corresponding regions of interest used in analysis (eyes, head, body, foreground objects, background objects).
Figure 5.2. Area-normalized fixation proportions for the People scenes as a function of Instruction (Told, Not Told), Session (Pretest, Test), and Region (eyes, heads, bodies, foreground objects, background). The Told group fixated the eyes more, and the heads less, than the Not Told group.
Figure 5.3. Area-normalized fixation proportions for the No People scenes as a function of Instruction (Told, Not Told), Session (Pretest, Test), and Region (foreground objects, background).
Figure 5.4. Proportion of first fixations landing in each region of the People scenes as a function of Instruction and Session.
References


Birmingham, E., Bischof, W.F., & Kingstone, A. Is there a default bias to select the eyes? Under review at *Quarterly Journal of Experimental Psychology*.


CHAPTER 6

Saliency does not account for fixations to eyes within social scenes

A version of this chapter is under review at *Vision Research*. Birmingham, E., Bischof,W.F., & Kingstone, A. Saliency does not account for fixations to eyes within social scenes.
Recent research has shown that observers have a fundamental bias to look at the eyes of people when they are presented within complex social scenes (Birmingham et al., 2008 (a), (b); under review (a),(b)). This preference for eyes is expressed early on, within the first two fixations (Birmingham et al. 2008b; under review (a),(b)), and persists across a variety of scenes and tasks. Furthermore, the bias to select eyes is enhanced by instructions to study social aspects of the scene (e.g., the attentional, emotional, or cognitive states of people in the scene). Furthermore, observers look overall at the eyes more frequently when presented with scenes with high social content than scenes with low social content (Birmingham et al., 2008 (a).(b)). When taken together, the extant research suggests that observers have a default bias to inspect the eyes of others because they understand them to be socially communicative stimuli that provide important information about a social scene (Birmingham et al., 2008b; under review (a),(b)).

One might ask, however, to what extent basic low-level visual factors, such as saliency (Itti, Koch, & Niebur, 1998; Koch & Ullman, 1985), contribute to the preferential selection of eyes? In other words, how much of the default bias toward eyes is simply reflecting the fact that eyes are highly conspicuous scene regions that attract fixations based on low-level visual features?

Bottom-up models of scene perception consider visual saliency to be a strong predictor of where observers will allocate their attention within natural scenes. The most popular saliency model (Koch & Ullman, 1985; Itti et al., 1998; Itti & Koch, 2000) is based on feature integration theory of visual search (Triesman & Gelade, 1980). The basic assumption of the saliency model is that before attention is focused on any one aspect of the scene, the visual field is processed rapidly and in parallel for basic visual features. The output of this ‘preattentive’ processing of the scene is the construction of several topographic feature maps, each coding for local differences in a particular feature (e.g., changes in intensity, color (red-green; blue-yellow), and edge orientation). These feature maps are computed across several spatial scales, varying from fine to course. Feature maps are then combined across scales into three “conspicuity maps”,

...
The three conspicuity maps are normalized to a fixed range (e.g., 0-1) and summed into the final saliency map, which is a modality-independent map coding for conspicuous (i.e., salient) scene locations. It is this saliency map which guides the deployment of attention. According to the ‘winner-take-all’ hypothesis, the most salient location in the map ‘wins’ focal attention (see Figure 6.1). After the winner is attended, attention moves along the remaining salient locations in order of decreasing saliency.

-- Figure 6.1--

While strong versions of saliency models claim that bottom-up saliency alone accounts for fixation behaviour (Itti et al., 1998), more flexible versions allow for modulation by top-down factors such as pre-knowledge of a search target to be found, task instructions, and scene semantics or gist (e.g., Torralba, 2003; Tsotsos et al., 1995; Wolfe, 1994). In addition, it is often assumed that the influence of top-down factors increases over time, such that initially attention is guided in mostly a bottom-up fashion by the saliency map, but is later under cognitive control (Parkhurst et al., 2002). In scene perception, this means that the first fixation (or possibly the first few fixations) is dominated by visual saliency, but later fixations are allocated according to scene semantics or internal goals (e.g., Henderson, Weeks, & Hollingworth, 1999; Parkhurst et al., 2002). Consistent with this claim, Parkhurst et al. (2002) found that saliency best predicted fixation position early on in viewing, and less so as viewing time proceeded.

Given that these models predict salient regions to be fixated more frequently than non-salient regions, particularly within the first few fixations of viewing the scene, one might question Birmingham et al.’s interpretation that fixations to eyes, at least those that occur initially, are due to participants’ understanding that eyes are socially important. That is, a potential explanation

\footnote{Other models propose early influence of both bottom-up saliency and top-down factors on the control of attention (e.g., Torralba, 2003). All of these more flexible saliency models, however, assume that an initial saliency map is computed, and that top-down ‘maps’ simply combine with the visual saliency maps to control the allocation of attention (although see Rao, Zelinsky, Hayhoe, & Ballard, 2002, for a different approach).}
for why there is a default bias to select the eyes is that eyes are very visually salient. What is needed is an analysis of how much visual saliency accounts for fixations committed to the eyes. We chose to conduct this analysis in three different ways.

**Basic performance of the saliency model.** First, we were interested in how well the classic saliency model of Itti and Koch (2000) accounted for the first fixation performance across a variety of experimental data (Birmingham et al., 2008 (a), (b), under review (a) (b)). We chose to analyze the first fixation data because saliency is predicted to be most influential early on in scene viewing, and so analyzing later (or all) fixations might underestimate the role of saliency in determining fixation position. For each experiment, we computed the average saliency of fixated scene locations and compared this value to two control values. The first control value was the average saliency of random locations sampled uniformly from the image (called "uniform-random"). To control for the known bias to fixate the lower central regions of scenes, the second control value was the average saliency of random locations sampled from the smoothed probability distribution of all first-fixation locations from participants’ eye movement data across all scenes (called "biased random"). These comparisons allowed us to determine whether the saliency model accounted for first fixation position above what would be expected by chance.

**Threshold analysis.** Second, we chose to represent the fixation data in a slightly different way to provide a ‘reality check’ on the importance of saliency in determining fixation position. Presumably, saliency values at fixated position could be quite low (e.g., 0.07), and yet be statistically higher than a random sampling of fixation positions (e.g., 0.01). This is in large part due to the sparseness of the saliency maps generated by Itti and Koch’s (2000) model (e.g., see Figure 6.2), which produce only a few salient regions, leaving the rest of the map black (i.e., non-salient). Itti (2005) has recently compared human fixations against a more appropriate threshold value for the saliency model, one that is higher than that which is expected by random sampling of the saliency maps (e.g., 0.01) but lower than what would be expected if saliency alone drove performance (1.0). In keeping with this approach we chose an
arbitrary saliency threshold value to indicate whether a fixation was on a salient region (regions with saliency values less than that threshold essentially appear black in the maps). In Itti (2005) the threshold saliency used was 0.25, but in our maps the threshold saliency value happened to be an even more liberal value of 0.10 (i.e., regions with saliency below 0.10 appear black).

-- Figure 6.2--

**Average saliency of regions.** Finally, we computed the relative saliency of the eyes versus the other regions in the scene. We know from the studies of Birmingham et al. that the eyes and heads are both highly likely to be fixated within the first one to two fixations (Birmingham et al., 2008b; Birmingham et al., under review (a),(b)), and more so than any other region (e.g., bodies, foreground objects). But how salient are the eyes and heads relative to the rest of the scene? Using saliency maps from Itti & Koch’s (2000) model, we compared the average saliency of eyes, heads, bodies, foreground objects, and background. This allowed us to determine whether the eyes or heads were more salient than the other regions, which might explain why they attracted initial fixations.

**Methods**

**Participants**

The fixation data from the studies reported in Chapters 2 – 5 were analyzed, giving a total of 119 participants (Birmingham et al., 2008 (a)(b); under review (a)(b)).

**Apparatus**

Eye movement data were originally collected using an Eyelink II tracking system. The on-line saccade detector of the eye tracker was set to detect saccades with an amplitude of at least 0.5°, using an acceleration threshold of 9500°/s² and a velocity threshold of 30°/s.

**Stimuli**

The images were full color digital photos taken of different rooms in the UBC Psychology building. Image size was 36.5 x 27.5 (cm) corresponding to 40.1° x 30.8° at the viewing distance of 50 cm, and image resolution was 800 x 600 pixels. Images in all experiments were selected from a pool of 90 scenes. In Experiment 1 (Birmingham et al., 2008a) observers saw a
set of 40 images, all containing either one or three people. In Experiment 2 (Birmingham et al., 2008b), a subset (20) of these images were shown. In Experiment 3 (Birmingham et al., under review (a)), 12 images were shown to observers, again all containing one or three people. In Experiment 4 (Birmingham et al., under review (b)), 15 images were shown in an initial study session, 12 of which contained one or more people, and 3 of which contained only objects. These plus an additional 41 images (total of 56) were shown in a later test session. In the test session, 32 of the 56 scenes contained one or more people, 16 scenes contained only objects, and 8 were filler scenes.

Itti and Koch (2000) have developed an algorithm that enables the measurement of the visual saliency of an image by identifying strong changes in intensity, colour and local orientation. The software is provided by Laurent Itti at [http://ilab.usc.edu/toolkit/downloads.shtml] and the model is described in detail by Itti and Koch (2000). The program’s default weightings were used. Examples of scenes and their corresponding saliency maps are shown in Figure 6.2. Saliency values were normalized to a range of 0 (absolutely non-salient) to 1 (highly salient).

Results

**Basic performance of saliency model.**

The saliency at the location of the first fixation was compared to the two chance-based estimates described above (*uniform-random* and *biased-random*). These data are presented in Table 6.1, broken down by experiment. Also, see Figure 6.2 for fixations overlaid on the saliency maps of each example image. To determine whether the saliency model accounted for first fixation position above what would be expected by chance, non-parametric statistics (Mann-Whitney U tests) were performed to compare the medians of *fixated* saliency and *uniform-random* saliency, and on the medians of *fixated* saliency and *biased-random* saliency.

--Table 6.1--

**Experiment 1** (Birmingham et al., 2008a). The fixated saliency was very low (0.011), as was uniform-random saliency (in fact, there were identical, 0.011, p>0.10). Fixated saliency
was also no different from biased-random saliency (0.017; p>0.10). Thus, the results of Experiment 1 indicate that the saliency at fixated locations was no higher than would be expected by the random models.

**Experiment 2** (Birmingham et al., 2008b). The fixated saliency was again very low (0.004), and no different than uniform-random saliency (0.011; p>0.10). Fixated saliency was also no different from biased-random saliency (0.008; p>0.10).

**Experiment 3** (Birmingham et al., under review (a)). The fixated saliency (0.029) was not statistically different from the uniform-random saliency (0.010), p>0.10, or from the biased-random saliency (0.005), p>0.10.

**Experiment 4** (Birmingham et al., under review (b)). People scenes: The fixated saliency (0.002) was actually significantly lower than uniform-random saliency (0.012), p<0.01. Fixated saliency was also significantly lower than biased-random saliency (0.017), p<0.0001. Thus, in Experiment 3 observers fixated regions that were less salient than would be expected by chance. No People scenes: The fixated saliency (0.023) was not statistically different from the uniform-random saliency (0.010), p>0.10, or from the biased-random saliency (0.005), p>0.10.

**Threshold analysis.**

The previous analysis revealed that both fixated and randomly generated saliency values were extremely low (close to zero) and did not differ statistically. Thus, the saliency model failed to predict the location of first fixation across four experiments. As a next step, we determined whether saliency at fixated locations was below an arbitrary saliency threshold. Recall that Itti (2005) chose an arbitrary threshold of 0.25 to indicate reasonable performance of the model, i.e., to indicate that a fixation had landed on a location that the model considered salient (a region with less than 0.25 was considered non-salient and was essentially black in their maps). In our maps, which were generated by the same model, the threshold was lower, i.e., a region with saliency less than 0.10 effectively appears black (see Figure 6.2 for examples). This most likely represents the fact that our images had a larger number of
reasonably salient regions than those of Itti (2005). Thus, for the threshold analysis we compared median saliency at fixated locations against 0.10.

For all experiments, median fixated saliency (Table 6.1) was significantly below the 0.10 threshold (Experiment 1: p<0.01; Experiment 2: p<0.01; Experiment 3: p<0.00001; Experiment 4: p<0.00001, confirmed by non-parametric comparisons). Thus, observers tended to fixate non-salient locations in the scenes.

**Average saliency of regions:**

Using saliency maps from Itti and Koch’s (2000) model, we compared the saliency of eyes, heads, bodies, foreground objects, and background. This allowed us to determine whether the eyes or heads were more salient than the other regions, which might explain why they attracted initial fixations in the studies by Birmingham et al. (2008b; under review (a) (b)). Although this interpretation is unlikely given that saliency did not account for first fixation behaviour, we wanted to be sure that eyes and heads were not more salient than other regions.

We computed the saliency of each region across all scenes used in the four experiments. The regions had been defined previously using Eyelink DataViewer (SR Research LTD; see Birmingham et al., 2008a). We immediately noticed that median saliency of eyes (0.012) and heads (0.049) were below the 0.10 threshold, i.e., these regions were not salient. This was also true of the other regions (foreground objects: 0.011; bodies: 0.024; background: 0.005). Note that some regions were quite large (e.g., background), meaning that even if parts of that region were highly salient, computing saliency over all of the region’s pixels would reduce its overall saliency value. However, this was not a problem for eyes or heads because they were relatively small.

A Kruskal-Wallis one-way ANOVA on ranks with Region as factor revealed that at least two medians were different (p<0.05). Multiple comparisons (Kruskal-Wallis z-value test) revealed that the eyes were no more salient than any other region, p>0.05. Foreground objects also did not differ from any region, p>0.05. The heads and bodies were statistically more salient than the background, p<0.05, although note that the median saliency values of most regions
were significantly below the 0.10 threshold. (eyes, p<0.01; bodies, p<0.0001; foreground objects, p<0.0001; background, p<0.00001), except for heads (0.049), which were only marginally below the threshold (p>0.07). By definition the fact that the saliency values were below threshold means that saliency cannot account for the fixation performance.

Discussion

The present study asked the following question: Is the initial default bias to select the eyes due (at least in part) to the eyes being visually salient? Previous research has shown that when presented with a complex scene containing people, observers are more likely to direct their first fixations to eyes and heads than to any other region (Birmingham et al., 2008 (a) (b); under review (a)(b)). Over time, observers show a clear preference for eyes over any other region, including heads. We had speculated that these findings reflect observers’ understanding that the eyes are socially communicative stimuli that provide important information about the meaning of a scene. However the extent to which bottom-up saliency alone could be driving observers fixation toward the eyes was unknown. The present study addressed this issue.

We used three analyses to assess how well saliency accounts for the data from four experiments, and found that saliency accounted for almost nothing. Not only did the saliency model do no better at predicting first fixations than would be expected by a random model, we also found that saliency at fixated locations was extremely low. In fact, in one experiment (Experiment 4, People scenes) observers fixated regions that were actually less salient than would be expected by chance. In addition, in all experiments saliency at the first fixation was significantly lower than the 0.10 threshold, i.e., observers tended to fixate parts of the scene that were black in the saliency maps. Furthermore, the eyes and heads were generally non-salient (median saliency close to 0, and either significantly below the 0.10 threshold (eyes) or marginally so (heads)). Thus, it is clear that saliency could not have been what drove observers to direct their early fixations to the eyes and heads of people in the scene.

That the saliency model did not do a good job at predicting first fixation locations is perhaps not entirely surprising. There is mounting evidence that saliency models are very poor
at predicting human fixations when the task is active, such as when walking down a hallway towards a target (Turano et al., 2003), making tea or sandwiches (Land & Hayhoe, 2001), or searching for items within a scene (e.g., Henderson, Brockmole, Castelhano, and Mack, 2007; Underwood & Foulsham, 2006; Zelinsky et al., 2006). Saliency also fails to predict human fixations when there are strong contextual effects (e.g., Torralba et al., 2006), and when the scene is a meaningful real world image (Henderson et al., 2007). Indeed, the model appears to perform above chance only when the task is unstructured and the scene is simplified, non-meaningful, and contains many highly salient regions (e.g., the fractals of Parkhurst et al., 2002), or when the scene contains salient motion cues (Itti, 2005). In the instances where the saliency model fails to predict human attention, it must be argued that either the implementation of the saliency model needs modification, that top-down guidance can largely override bottom-up control of attention, or that saliency simply does not play a meaningful role in guiding attention in active tasks involving rich (often real-world) stimuli. As we have shown, saliency does a particularly poor job at predicting fixations within real world scenes containing people.

What does guide fixations in scenes with people? A recent paper by Cerf et al. (2008) showed that the model that best predicted where observers committed fixations within scenes with people (for 2 seconds of viewing) was a saliency model combined with a face-detection model. This combined model outperformed the saliency model on its own, which performed just above chance. Indeed, even the face-detection model on its own performed better than the saliency model on 83/143 images, but this fraction did not reach significance. An analysis of participants’ fixation data showed that on 88.9% of trials, a face was fixated after two fixations. This finding supports our findings that observers are highly likely to fixate the heads and eyes within the first two fixations (Birmingham et al., 2008b, under review (a),(b)). When taken together, the extant research suggests that humans are highly sensitive to the presence of people, particularly the information from faces and eyes. Saliency, on the other hand, appears to play a relatively trivial role in this process.
Having determined that saliency is not driving observers to make early fixations to the eyes of people in a scene, then what is? At the beginning of this paper we suggested one interpretation: that observers look to the eyes because they understand them to be socially communicative stimuli. This interpretation is bolstered by the finding that fixations to eyes are enhanced by tasks instructing observers to study social aspects of the scene, such as the attentional, emotional and cognitive states of people in the scene (Birmingham et al., 2008b; Birmingham et al., under review (a)). Note that our instructions never explicitly told people to look at the eyes of people in the scene, and so these results suggest that observers share an understanding that the eyes provide important information about the mental states of others. That people select eyes regardless of task may suggest that they are naturally interested in the social information provided by eyes.

This default interest in the social information from other people’s eyes may be a hallmark of normal social cognition, one that develops early in life. Indeed, even young infants are highly sensitive to the presence of other people’s eyes, and begin to preferentially scan the eyes of a face by about 2 months of age (Haith, Bergman & Moore, 1979; Maurer & Salapatek, 1976). Baron-Cohen (1994) has even suggested that the preference for the eyes is supported by a specific module, called the Eye Detection Director (EDD), which both detects the presence of eyes and computes the direction of gaze. It is also thought that a failure to develop an interest in the social information from other people’s eyes is an underlying factor in social disorders like autism (Baron-Cohen, 1994; Dawson et al., 1998; Klin et al., 2002). Thus, there is accruing evidence that, as part of normal social development, humans have a fundamental tendency to rapidly select and process the social information from eyes.
Table 6.1. Median values for saliency of fixated regions, uniform-random saliency, and biased-random saliency, as a function of experiment. In all experiments, the saliency of fixated locations was no different from uniform-random saliency or biased-random saliency (ps>0.10), indicating that the saliency model did not predict the location of first fixations.

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Median saliency of fixated location</th>
<th>Uniform-Random saliency</th>
<th>Biased-random saliency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.011</td>
<td>0.011</td>
<td>0.017</td>
</tr>
<tr>
<td>2</td>
<td>0.004</td>
<td>0.011</td>
<td>0.008</td>
</tr>
<tr>
<td>3</td>
<td>0.029</td>
<td>0.010</td>
<td>0.005</td>
</tr>
<tr>
<td>4</td>
<td>0.002</td>
<td>0.012</td>
<td>0.017</td>
</tr>
</tbody>
</table>
Figure 6.1. The original saliency model, adapted from Koch and Ullman (1985). Visual input is processed rapidly and in parallel for basic visual features like colour, intensity, and orientation. Separate topographic feature maps are computed across several spatial scales, each coding for local differences in a particular feature (e.g., changes in intensity, color (red-green; blue-yellow), and edge orientation). Feature maps are then combined across scales into three “conspicuity maps”, one for intensity, one for colour, and one for orientation. The three conspicuity maps are normalized to a fixed range (e.g., 0-1) and summed into the final saliency map, which is a code for conspicuous (i.e., salient) scene locations. According to the ‘winner-take-all’ hypothesis, the most salient location in the map ‘wins’ focal attention. After the winner is attended, attention moves along the remaining salient locations in order of decreasing saliency.
Figure 6.2. Examples of the scenes used, and their corresponding saliency maps (Itti & Koch, 2000) overlaid with the first fixations of Experiment 1.
References


Birmingham, E., Bischof, W.F., & Kingstone, A. (under review (a)). Is there a default bias to select the eyes? Under review at *Quarterly Journal of Experimental Psychology*.

Birmingham, E., Bischof, W.F., & Kingstone, A. (under review (b)). Saliency does not account for fixations to eyes within social scenes. Under review at *Vision Research*.


CHAPTER 7

Get real! Resolving the debate about equivalent social stimuli.

A version of this chapter is under review at *Psychological Science*. Birmingham, E., Bischof, W.F. & Kingstone, A. Get real! Resolving the debate about equivalent social stimuli.
Our everyday knowledge suggests that we are very interested in the attention of other people. Indeed, experience suggests that as social beings we are quick to notice when people are looking at us, and when they are not looking at us we are quick to see what they are looking at. This intuition, that we care about where other people are attending, has led to the birth of research in social attention.

While there are several cues to the direction of another person’s attention (e.g., gaze direction, head position, body position, pointing gestures), the above description suggests that gaze direction has a special status as an attentional cue (Emery, 2000; Langton et al., 2000). Morphologically, the human eye is equipped to promote fast discrimination of gaze direction, having the highest iris-to-sclera contrast of all the primate eyes (Kobayashi & Koshima, 1997). Humans are not only very accurate at discriminating gaze direction (Cline, 1967; Gibson & Pick, 1963; Lord & Haith, 1974), but we also appear to have neural structures that are preferentially biased for processing gaze information. For instance, single cell recordings in monkeys show that the superior temporal sulcus (STS) has cells that are selective for different gaze directions, independent of face orientation (Perrett et al., 1985); and neuroimaging studies (Hoffman & Haxby, 2000; Pelphrey et al., 2004) have similarly shown that the human STS seems to be especially activated by changes in gaze direction. Indeed, eye gaze is thought to be so important that it has been placed as the primary social attention cue in prominent models of social attention (Baron-Cohen, 1994; Perrett et al., 1992).

Perhaps what makes eyes so unique is that in addition to showing someone’s direction of attention, they can be used to infer a wealth of other social information that we use on an everyday basis. For instance, the eyes can help us determine what someone is feeling, thinking, or wanting (Baron-Cohen et al., 1997). The eyes are also used to affect social interactions, by facilitating conversation turn-taking, exerting social dominance, or signalling social defeat or appeasement (Argyle & Cook, 1976; Dovidio & Ellyson, 1982; Ellsworth, 1975; Exline, 1971; Exline et al., 1975; Kendon, 1967; Kleinke, 1986; Lochman & Allen, 1981). Thus, both intuition
and empirical evidence suggest that the eyes are extremely important and unique social-communicative stimuli.

To capture the unique social importance of eyes, an abundance of research has examined the extent to which gaze direction can trigger an attention shift in others (or ‘joint attention’). It is now well-established that infants (Hood, Willen & Driver, 1998; Farroni et al., 2000), preschool children (Ristic, Friesen & Kingstone, 2002) and adults alike (Driver et al., 1999; Friesen & Kingstone, 1998; Langton & Bruce, 1999) shift attention automatically to where others are looking. In the ‘gaze cueing’ paradigm used to test this, a participant is first shown a picture of a real or schematic face with the eyes looking either to the left or to the right. A target is then presented either at the gazed-at location or at the nongazed-at location. The typical finding is that response time (RT) to detect a target is faster when the target appears at the gazed-at (cued) location than at the nongazed-at (uncued) location, indicating that attention has been allocated to the gazed-at location. Because this effect emerges rapidly and occurs even when gaze direction does not predict where a target is going to appear, it is considered to indicate reflexive orienting of attention.

The reason this discovery has been so exciting to attention researchers is that for decades it was thought that central non-predictive cues do not produce reflexive orienting (Jonides, 1981). That is, it was always assumed that a central cue (an arrow, for instance) had to correctly predict the location of the target on the majority of trials in order for it to induce orienting, and that this orienting was thus voluntary in nature. The discovery that orienting can be induced by a central gaze cue that does not predict the location of a target suggested that eyes are unique, socially and biologically important stimuli that can induce reflexive orienting in a manner unlike any other central cue (Friesen & Kingstone, 1998).

But is it the case that gaze cues are truly special in this regard? More recent studies have shown that reflexive shifts in attention are not unique to central gaze cues. In particular, several studies have now shown that non-predictive central arrow cues also produce reflexive orienting effects. In fact, arrow cueing effects are largely indistinguishable from the gaze cueing
effect (Hommel, Pratt, Colzato, & Godijn, 2001; Pratt & Hommel, 2003; Ristic, Friesen, & Kingstone, 2002; Ristic, Wright & Kingstone, 2007; Tipples, 2002; see also Eimer, 1997).

Indeed, even some of the more subtle behavioural signatures of orienting that were initially thought to be unique to gaze cues have now been found for arrow cues. For instance, it was initially thought that only gaze cues produce reflexive orienting despite observers’ intentional orienting to other locations, otherwise known as ‘counterpredictive’ cueing (Driver et al., 1999; Friesen et al., 2004). However, Tipples (2008) found that arrows, too, produce counterpredictive cueing, and reported reflexive orienting to the location pointed at by both arrows and gaze cues even when a target was nearly ten times as likely to appear elsewhere.

Furthermore, although some neuroimaging studies suggest that gaze and arrow cueing may be subserved by different neural systems (Kingstone et al., 2000; Ristic et al., 2002; Experiment 3; Vuilleumier, 2002), other studies conflict with this claim (Tipper et al., 2008; Hietanen et al., 2006). For instance, Tipper et al. (2008) studied gaze and arrow cueing using an ambiguous stimulus that could be perceived as either an eye or an arrow, thus removing the physical stimulus differences normally present in comparisons of gaze and arrow cueing (e.g., Hietanen et al., 2006)

2. Tipper et al. found very few differences between the neural activations underlying gaze and arrow cueing, save for a bigger sensory gain at the cued location for gaze cues than for

\[ \text{\underline{\text{\footnotesize{\text{\textsuperscript{2}}}}}} \]

Indeed, it may be the case that the brain activation differences produced for gaze and arrow cues may be partly due to the recruitment of different brain areas for visually analyzing gaze and arrow cues, and not necessarily for the subsequent shifts of attention (Hietanen et al., 2006). For instance, it is known that STS is more highly activated for visual discriminations of gaze direction than for discriminations of arrow direction (e.g., Hooker et al., 2003). Once these basic visual processing differences are removed, and only the subsequent orienting of attention is examined, it is less obvious that gaze and arrow cues are subserved by distinct attentional mechanisms (Hietanen et al., 2006; Tipper et al., 2008). For instance, Hietanen et al. (2006) found that reflexive orienting was supported by partially different mechanisms for gaze cues and arrow cues, with the main difference being that orienting in response to arrow cues activated a more distributed network than orienting in response to gaze cues, and that some of these additional brain areas for arrow cueing (e.g., frontal eye fields) are thought to be involved in voluntary orienting. This might suggest that gaze cueing is slightly more reflexive than arrow cueing, even though behaviourally this distinction is not apparent. However, note that Tipper et al. (2008) did not find this neural distinction between gaze and arrow orienting when the stimulus differences between gaze and arrow cues were removed.
arrow cues. Thus, the authors note that their findings support the idea that gaze cues do not engage different neural mechanisms for orienting attention, but that gaze cues may instead engage the same neural circuits somewhat more vigorously than arrow cues. When taken together, there is only equivocal evidence to suggest that gaze cueing is unique; pointing to the general conclusion that gaze and arrow cues are treated similarly by the human attention system.

One interpretation of these findings is that arrows and eyes are equally socially relevant cues, and are thus given equal priority by the attentional system. Although contrary to the general intuition that eyes are unique social stimuli, it is clear that arrows can, like eyes, communicate directional information that may be prioritized by the human attention system. This value of arrows has not been overlooked (Kingstone et al., 2003; Tipples, 2002). For instance, Kingstone et al. (2003) note that “arrows are obviously very directional in nature, and, like eyes, they have a great deal of social significance. Indeed, it is a challenge to move through one’s day without encountering any number of arrows on signs and postings” (p. 178). Indeed, an arrow is a prime example of a communicative symbol that humans use to direct other people’s attention to particular parts of a shared environment (Tomasello & Call, 1997). Thus, upon reflection, it is perhaps not entirely counter-intuitive that arrows, like eyes, produce reflexive shifts in attention.

A second, somewhat more radical interpretation of the finding that eyes and arrows are equivalent within the gaze cueing paradigm is that this paradigm may not be capturing key aspects about eyes that distinguish them as unique social stimuli. According to this view, the general intuition that eyes are special and distinct from arrows is valid, but the cueing paradigm is not measuring what makes eyes and arrows different. Rather, the cueing paradigm is measuring eyes and arrows on a dimension that they share a great deal of similarity, i.e., their ability to communicate directional information (Gibson & Kingstone, 2006). An analogy would be taking a 150 pound person and a 150 pound rock, weighing them, and concluding that they are the same. They are the same, in terms of weight, but clearly we have the intuition that they
are not the same in many other ways. For instance, we suspect that for various reasons people would, in general, be much more interested in the person than the rock. To demonstrate that, however, one needs a different measurement, i.e., a different research approach. In much the same way, what may be needed for eyes and arrows is a different approach -- one that might better reflect our intuition that the human attention system cares about eyes in a way that it does not care about other stimuli in the environment, including arrows.

An alternative approach may be found by considering the different components of attention that can be measured in experiments involving eye and arrow stimuli. Importantly, both gaze and arrow cueing studies are specifically designed to test one particular component of visual attention: spatial orienting. However, as we will describe later, spatial orienting in response to a cue is only one component of attention. Another component of attention, selection, is not captured by these cueing studies. Consider a real world example of social attention: You are walking across campus and notice that your colleague is standing on the sidewalk and looking at something on the ground. Using her gaze direction you orient your attention to see what she is looking at. It is clear from this example that there are at least two distinct components of attention: first, you select your colleague’s eyes as a key social stimulus, and second, you shift your attention from her eyes to the location/object that she is looking at. The selection component of attention is relatively trivial within the context of the cueing paradigm because the cue, that is, a gaze or arrow stimulus, is presented at central fixation and typically in advance of the target object (Gibson & Kingstone, 2006). In other words, it is not known whether the selection process differs for eyes and arrows, because this is not tested in the cueing paradigm. If this critical component of attention differs for eyes and arrows, it would shed light on whether or not eyes are distinct from arrows and other environmental stimuli, insofar as eyes are given priority by the attentional system.

The fact that no studies have compared the selection of eyes versus arrows or other objects in the environment is noteworthy given the strong tradition of research on selective attention (e.g., James, 1890; Broadbent, 1958; 1972; Deutsch & Deutsch, 1963; Moray, 1959;
Neisser, 1967; Treisman, 1960). The basic assumption behind all these conceptualizations of selective attention is that humans possess a capacity limitation when it comes to handling information in the world. The implication of this capacity limitation is that we must select some items for processing at the expense of others (hence the term *selective* attention). Space-based theories claim that selective attention is inherently spatial in nature, e.g., *locations* in space are selected. Furthermore, there are two components of attention: *selection*, whereby some items are chosen for further processing and others are ignored; and *spatial orienting*, whereby focal attention is shifted from one spatial location to another (Theeuwes, 1993). These components are intricately linked: shifting focused spatial attention to a location is thought to be the operation by which information in the environment is selected for further processing. How does this attentional focusing occur? One popular metaphor is that attention is like a spotlight shining on selected locations, with items within the spotlight benefiting from enhanced perceptual processing (e.g., Broadbent, 1982; Downing & Pinker 1985; Eriksen & Hoffman, 1973; Posner, 1980: Shulman et al., 1979; Tsal 1983). This spotlight is limited in capacity, that is, in its spatial extent (Yantis, 1988). Thus, this spotlight must *shift* in order to attend to items at different locations in the display (spatial orienting).

Gaze and arrow cueing studies focus exclusively on how gaze and arrow cues affect the spatial orienting component of attention. There are two standard methods for ensuring maximum comparability between gaze and arrow cues. First, the cues are matched as closely as possible, so that the only aspect of the cue that changes is its direction. Second, in order to ensure that cue direction is indeed processed (even if irrelevant to the task, i.e., the cue is spatially noninformative), the cue is presented at the current focus of attention, i.e., at the central fixation point. Because the cue is presented within the current focus of attention, it is inevitably selected. As a consequence of being selected, the cue is identified (e.g., arrow) and its direction (e.g., left or right) is processed. As mentioned earlier, central arrow and gaze cues produce reflexive shifts of attention. Thus, as a consequence of processing the central cue, spatial attention is shifted obligatorily from the central cue to the pointed-at (cued) location.
Because cuing studies only examine spatial orienting in response to central gaze and arrow cues, these studies do not inform us about the initial selection of gaze and arrow cues from the environment. Indeed, because cueing studies are solely interested in the orienting aspect of attention, the experimenter pre-selects the cue and places it at fixation (the current focus of attention). When the selection process of attention is omitted, the prevailing literature indicates that eyes and arrows are given equal priority by the attention system. Does this equivalence hold, however, when the selection component of attention is measured? In other words, will eyes and arrows be given equal priority when participants are provided with the opportunity to select stimuli? The aim of the present investigation was to address this issue. We did this by presenting gaze and arrows within complex scenes and studying what people selected to fixate.

Research on scene perception has consistently shown that when presented with a complex scene, observers tend to select (fixate) items they find interesting or informative (Buswell, 1935; Henderson et al., 1999; Loftus & Mackworth, 1978; Yarbus, 1967). When people are absent from a scene, this means that observers will look primarily at objects that add semantic meaning to the scene and scene regions with high amounts of visual information (Antes, 1974; Buswell, 1935; Henderson et al., 1999; Loftus & Mackworth, 1978). When people are present in a scene, observers look primarily at the eyes and faces of the people and spend much less time looking at the rest of the scene (Birmingham et al., 2008 (a),(b), under review (a),(b); Yarbus, 1967). This suggests that when a person is presented in a scene, their eyes become important to understanding the scene (Birmingham et al., 2008 (a),(b) under review (a), (b)). Indeed, we have interpreted these findings as indicating that people fixate the eyes of others because they perceive the eyes to contain important social information. In support of this, observers’ preference for eyes is enhanced by social tasks (e.g., describe where people in the scene are directing their attention) and by increasing the social content and activity of a scene (e.g., increasing the number of people actively doing something); (Birmingham et al., 2008 (a),(b)). Thus there is evidence suggesting that observers preferentially select gaze
information from a complex scene, and that this reflects the fact that eyes are perceived to be informative social stimuli.

However, no studies have tested whether gaze would be preferentially selected if an arrow were also placed in the scene. Research using the cuing paradigm indicates that eyes and arrows are equivalent attentional cues, and that this reflects the fact that arrows, like eyes, are socially significant: “arrows are obviously very directional in nature, and, like eyes, they have a great deal of social significance. Indeed, it is a challenge to move through one’s day without encountering any number of arrows on signs and postings” (Kingstone et al., 2003, p. 178). One possible outcome then, as predicted by the extant data and theory, is that eyes and arrows will be selected to the same extent. An alternative possibility is that eyes will be preferentially selected over arrows. This finding would suggest that eyes and arrows do not have equal social relevance, and would dovetail with the general intuition that while eyes and arrows are both directional, eyes are unique in that they can communicate other social information, such as the emotions, intentions, states of mind, and ages of other people. As such, eyes may be prioritized by the attention system, as evidence by preferential selection, i.e., fixation. This outcome is not predicted by the gaze and arrow cueing studies.

Thus, the present study examined the extent to which eyes and arrows are selected from complex scenes. We presented a variety of photographs of real world scenes containing both people and arrows, and monitored observers’ eye movements while they freely viewed the scenes. This allowed us to determine how often, and how quickly, observers selected eyes and arrows.

Method

Participants

Fifteen undergraduate students from the University of British Columbia participated in this experiment. All had normal or corrected to normal vision, and were naïve to the purpose of the experiment. Each participant received course credit for participation in a one-hour session.
**Apparatus**

Eye movements were monitored using an Eyelink II tracking system. The on-line saccade detector of the eye tracker was set to detect saccades with an amplitude of at least 0.5°, using an acceleration threshold of 9500°/s² and a velocity threshold of 30°/s.

**Stimuli**

Full color photographs were collected from various sites on the World Wide Web. Each picture was presented on a white 800 × 600 pixel canvas. Thus, in some cases, a picture that was slightly smaller than 800 × 600 pixels was surrounded by the white borders of the canvas. Image (canvas) size was 36.5 x 27.5 (cm) corresponding to 40.1° x 30.8° at the viewing distance of 50 cm. Twenty three images were used in the present experiment: six images contained both people and arrows, one image contained arrows but no people, and sixteen remaining ‘filler’ images were displayed (containing photographs of people, faces, and paintings). The seven arrow images analyzed in the present experiment are shown in Figure 7.1.

--Figure 7.1--

**Procedure**

Participants were seated in a brightly lit room, and were placed in a chin rest so that they sat approximately 50 cm from the display computer screen. Participants were told that they would be shown several images, each one appearing for 15 seconds, and that they were to simply look at these images.

Before beginning the experiment, a calibration procedure was conducted. Participants were instructed to fixate a central black dot, and to follow this dot as it appeared randomly at nine different places on the screen. This calibration was then validated, a procedure that calculates the difference between the calibrated gaze position and target position and corrects for this error in future gaze position computations. After successful calibration and validation, the scene trials began.
At the beginning of each trial, a fixation point was displayed in the centre of the computer screen in order to correct for drift in gaze position. Participants were instructed to fixate this point and then press the spacebar to start a trial. One of 23 pictures was then shown in the center of the screen. Each picture was chosen at random and without replacement. The picture remained visible until 15 seconds had passed, after which the picture was replaced with the drift correction screen. This process repeated until all pictures had been viewed.

Results

Data handling

For each image, an outline was drawn around each region of interest (e.g., "eyes", "arrow") and each region’s pixel coordinates and area were recorded. We defined the following regions in this manner: eyes, heads, body (including arms, torso and legs), arrows, and ‘other’.

To determine what regions were of most interest to observers we computed fixation proportions by dividing the number of fixations for a region by the total number of fixations over the whole display. We corrected for area differences between regions and across scenes to control for the fact that large regions would, by chance alone, receive more fixations than small regions. This was accomplished by dividing the proportion score for each region by its area (Birmingham et al., 2008 (a), (b); Smilek et al., 2006).

To determine where observers’ initial saccades landed in the visual scene, we computed the number of first fixations that landed in a region (initial fixations). These data were not area-corrected.

Fixation proportions

Scenes with eyes and arrows: Our main question of interest was whether eyes and arrows would be fixated to the same extent. Thus, we analyzed the images containing both eyes and arrows, i.e., images with people who were large enough for the observer to see the eye region (Figure 7.1, A-C). Figure 7.2 (A-C) shows fixation plots for all subjects for these three images. Immediately noticeable from these plots is that observers concentrated their
fixations primarily on the people, particularly their eyes. In addition, observers rarely fixated the arrows.

---Figure 7.2---

To confirm these impressions, we conducted a Repeated Measures ANOVA on fixation proportions with Region (eyes, heads, bodies, text, arrows, and ‘other’ (the remainder of the scene)) as a factor. These data are shown in Figure 7.3. This analysis revealed a highly significant effect of Region ($F(5,70)=50.98$, $p<0.0001$). Pairwise comparisons (Tukey-Kramer multiple comparisons test) revealed that observers fixated the eyes (0.45) more than any other region ($p<0.05$). Heads and text were fixated the next most frequently (heads: 0.23; text: 0.21), and more so than bodies (0.05) and arrows (0.05), and the rest of the scene (0.01), $p$s<0.05. Importantly, it is clear that arrows were not fixated often in these scenes. Thus, to answer the main question of our study, eyes were fixated far more frequently than arrows, which were hardly fixated at all.

---Figure 7.3---

Scenes with larger arrows: One might wonder if observers failed to show a preference for arrows in the previous analysis because they were relatively inconspicuous compared to the rest of the scene (Figure 7.1, scenes A-C). While we chose arrows that were usually central and reasonably conspicuous, and that were of equal or greater size than the eyes, it is possible that the complexity of the scenes pulled attention away from the arrows. Thus, we analyzed three other scenes in which the arrows were very large, the scene was simple, and the people were small (Figure 7.1, D-F). Fixation plots for these images are shown in Figure 7.2 (D-F). Again, it is immediately noticeable that observers focused mostly on the people, particularly the heads of the people, and that few fixations were committed to the arrows. Even the empty bench in Scene E received more fixations than the arrows in the scene. For these scenes, we analyzed fixation proportions as a function of Region (heads, bodies, bench, arrows, other). The fixation proportion data are shown in Figure 7.4. Note that because the people were small, eyes were not visible and thus were not analyzed. The ANOVA revealed an effect of Region.
(F(4,56)=83.62, p<0.00001), with heads (0.51) being fixated more than any other region (Tukey Kramer, p<0.05). Bodies were the next most fixated (0.33), and more so than benches (0.09), arrows (0.05) and other (0.02), p<0.05. Thus, despite their very large size, arrows were again fixated infrequently relative to the people. (Note that as the data are proportions of fixations as a function of region, we could not directly compare the data for Scenes D-F to Scenes A-C, as the content of the scenes (and thus the regions of interest) was not constant.

--Figure 7.4--

Scenes with no people: The results thus far have demonstrated that observers care very little about arrows placed in a complex scene containing people. It appears that as social beings, observers allocate their attention primarily to other people, particularly their eyes. What happens when no people are in the scene? Given that arrows have been thought of as socially relevant objects (Kingstone et al., 2003; Tipples, 2002), would they receive preferential attention when placed among other objects (and with no people present)? We analyzed the data for the scene in Figure 7.1(G). For the analysis the scene was parsed into four regions: the bunch of grapes, noentry (the red ‘no–entry’ sign), arrow, and other (remainder of the scene). The fixation plot in Figure 7.2(G) shows that relative to the noentry sign and the grapes, the arrow was fixated infrequently. These data are summarized in Figure 7.5. An ANOVA on the fixation proportions revealed an effect of Region (F(3,42)=135.78, p<0.00001), with pairwise comparisons revealing that observers looked mostly at the noentry region (0.53) than any other region (Tukey-Kramer, p<0.05). The next most fixated region was ‘grapes’ (0.27), which was fixated more than the arrow (0.17), p<0.05. All three of these regions were fixated more often than the remainder of the image (other: 0.03), p<0.05. These data suggest that observers show little interest in the arrow relative to other objects. Note again that as the data are proportions of fixations as a function of region, we could not directly compare the data for Scene G to Scenes A- F, as the content of the scenes (and thus the regions of interest) changed.

--Figure 7.5--
Initial fixations

**Scenes with eyes and arrows:** Although the fixation proportions showed that eyes were fixated more frequently than arrows, these were averaged over the entire viewing period. Thus, the analyses of fixation proportions might reflect more voluntary or strategic viewing patterns that developed over time. The very first fixation, on the other hand, reveals which regions attract attention immediately upon the appearance of the scene. We reasoned that if arrows capture attention as strongly as eyes, then this would be reflected in the first fixation being just as likely to land on an arrow as an eye region. Thus, we analyzed the proportion of first fixations (the first fixation after the experimenter-controlled fixation at centre) that landed on eyes, heads, bodies, arrows, or other. These data were not area normalized. Figure 7.6 shows these data for scenes containing eyes and arrows. There was an effect of Region (F(5,70)=3.61, p<0.01), with eyes, heads, text, and the remainder of the scene all equally likely to receive the first fixation (eyes: 0.20; heads: 0.20; text: 0.22; other: 0.24), and all more likely than the arrow, which never received the first fixation (0.00).

--Figure 7.6--

**Scenes with larger arrows:** Larger arrows were also unlikely to receive the first fixation (Figure 7.7). An ANOVA revealed an effect of Region (heads, bodies, bench, arrow, other), F(4,56)=9.33, p<0.0001. Pairwise comparisons (Tukey-Kramer) revealed that bodies (0.40) and other (0.40) were both most likely to get the first fixation, more so than any other region (bench: 0.13; head: 0.07, arrow: 0.00), p<0.05. As with scenes containing smaller arrows, larger arrows never received the first fixation.

--Figure 7.7--

**Scenes with no people:** Are arrows fixated first when people are absent? Which regions of Scene G are most likely to be fixated first? An ANOVA revealed an effect of Region (noentry, grapes, arrow, other), (F(3,42)=70.38; p<0.0001). Pairwise comparisons revealed that the grapes were highly likely to be fixated first (0.93), and more so than the noentry sign (0.07), the arrow (0.00) or the rest of the scene (other: 0.00). These data are shown in Figure 7.8.
Discussion

The aim of present study was to determine whether eyes and arrows are selected to the same extent within complex scenes. While gaze cueing studies have found that directional gaze and arrow cues have similar effects on spatial orienting, these studies do not inform us whether the attentional system also selects eyes and arrows to the same extent. We reasoned that measures of selection might be more sensitive to the unique social importance of eyes, and as such might reveal an attentional (selection) priority for eyes over arrows.

One possibility was that observers might select arrows as often as eyes. This would be consistent with research using the cueing paradigm showing that gaze and arrows are equivalent attentional cues, suggesting that they are of equal social relevance. An alternative possibility was that eyes would be preferentially selected over arrows. This finding would suggest that eyes and arrows do not have equal social relevance, and would be in line with the general intuition that while eyes and arrows are equally good at conveying directional information, eyes are special social stimuli because, for example, they can communicate other important social information about people such as their age, identity, emotions, intentions, and so forth. As such, one would expect humans to prioritize information from eyes over arrows.

The results of the present study were clear. When both eyes and arrows were visible in a scene, the majority of fixations went to the eyes, and very few went to the arrows (Figure 7.3). Furthermore, an analysis of the first fixation made to the scene revealed that observers never fixated the arrow first (Figure 7.6). Instead, they were equally likely to fixate the eyes, heads, and text on the first fixation. This finding suggests that when presented in scenes with eyes, arrows do not capture attention, but eyes (and other regions) do. As the viewing session proceeded, observers showed a continuing interest in the people, particularly their eyes, and continued to largely ignore the arrows.

A general preference for people persisted in scenes in which the arrow was large and the people were small. We were interested in whether making the arrow large in comparison to
the people would successfully attract fixations and reduce interest in the people. Interestingly, however, in those images, the arrows were again rarely fixated (Figure 7.4). Instead, the heads were fixated the most frequently overall. Again, the arrows were never fixated first, and yet the bodies and remainder of the scene were both highly likely to be fixated first (Figure 7.7). Thus, even when arrows were large, they did not capture attention, nor did they receive many fixations overall relative to the people in the scene.

Finally, we were interested in whether an arrow would be preferentially selected when placed in a scene without people. Given that arrows have been thought of as social tools (Kingstone et al., 2003), one might expect them to receive more attention than other, presumably less social objects (as intuited, for instance, from the gaze/arrow cueing literature). Thus, we showed an image of a road sign with other graphic components (a bunch of grapes, a ‘no-entry’ symbol). However, the data revealed that observers again fixated the arrow less often than the other elements of the scene (Figure 7.5). In addition, the arrow was never fixated first, but the grapes often were (Figure 7.8). Thus, even when placed in a scene without people, arrows did not capture attention nor did they receive much attention over time.

There are several important implications of the present findings. First, to answer the main question of the study, when eyes and arrows are presented within complex scenes and observers are allowed to select items for further processing, observers show a preferential selection of eyes over arrows. This is consistent with the general intuition that eyes are special social stimuli that receive attentional priority. Thus, while eyes and arrows are equally good at conveying directional information, and thus produce equivalent effects on shifts of spatial attention within the cueing paradigm, they are not given equal priority by the attention system via selection within complex scenes. On the contrary, observers show a bias to select information from people’s eyes (Birmingham et al., under review (a)).

Second, arrows were not only selected less often than eyes, they were typically selected less often than most other scene regions. This was true even when the arrows were large, and when people were absent from the scene. What makes this finding interesting is that even
though it is clear that an arrow will produce reflexive shifts of attention within the context of the
cueing paradigm, observers show very little interest in arrows within the context of complex
scenes. One interpretation of these findings is that the importance of arrows as social
communicative tools may be restricted to situations in which direction or location information is
task relevant (e.g., following an exit sign on the highway; determining which lane is the turning
lane, etc.). Indeed, the cueing paradigm is just that – a situation in which the task is to detect a
target at a location on the screen. In that situation, even though arrow direction is not spatially
informative about where the target will appear, spatial location is a task relevant dimension, i.e.,
the only factor over which the target varies is its spatial position. In relation to the present study,
another future investigation is to determine whether arrows are selected more often from
complex scenes when the task is direction/location relevant (e.g., locate a target in the scene).
However, insofar as the task involves fixating different regions of a complex scene, one could
argue that location information is important and task relevant. Thus, even here, one finds that
arrows are not prioritized by the attention system.

A third implication of the present study is that despite a general preference to select
people from complex scenes, there appears to be a hierarchy to the selection of ‘people parts’.
If the people are large enough so that the eyes are visible, observers will concentrate their
fixations on the eyes, followed by the heads, and then bodies. If the people are too small for the
eyes to be discriminated, then observers will concentrate their fixations on the heads, followed
by bodies. Thus, while there is a general preference for people, observers preferentially fixate
the eyes if they are available. This is consistent with Perrett et al.’s (1992) model of social
attention, in which gaze is at the top of a hierarchy of cues to the direction of attention, followed
by head position and body position. The model claims that the eyes are used preferentially to
infer direction of attention, but when the eyes are unavailable people will rely on head position,
and when head position is unavailable they will rely on body position. Although we did not
specifically ask observers to infer the direction of attention of people in the scene, their fixation
patterns might reflect a natural hierarchy of social cues.
Now that we have shown a preferential selection of eyes over arrows, we return to our initial interpretation of the finding that eyes and arrows are equivalent within the context of the cuing paradigm. We ask the following question: is it problematic that the unique social importance of eyes is not detectable within the gaze cueing paradigm? As we mentioned before, one possibility is that the gaze cueing paradigm is simply not capturing key aspects about eyes that distinguish them as unique social stimuli. Certainly, if it were, researchers would have found that reflexive cueing is unique to central gaze cues, or that gaze cueing effects are substantially different from arrow cueing effects. Given that this has not been found, we proposed that the similarity between gaze and arrow cues within the cueing paradigm may be because this paradigm is not measuring what makes eyes and arrows different. Instead, we suggested that the cueing paradigm is measuring eyes and arrows on a dimension on which they have been equated for the purpose of the paradigm: their ability to convey directional information. In contrast, the present study showed that the unique social importance of eyes is reflected in measures of selection within complex scenes. What does this mean for social attention research?

On the one hand, the fact that eyes and arrows produce equivalent effects on shifts of attention in the cueing paradigm is not problematic at all. Certainly, cueing studies do not claim to elucidate whether gaze and arrows would be selected to the same extent in the real world – they simply show that once selected, gaze and arrows produce similar orienting effects. If one wants to know about selection, one must use a different research approach (such as that used in the present study). Furthermore, while the cueing paradigm does not seem to reflect the unique social quality of eyes relative to other cues, it does show that people will orient their attention reflexively to gazed-at locations, which is thought to be an important component of social attention (i.e., establishing joint attention). Thus, if one just considers the results with gaze cues, the cueing studies simply confirm that gaze direction is an effective attentional cue, something that was initially proposed by social attention models (e.g., Baron-Cohen, 1994; Perrett et al., 1992) but never tested until recently (Driver et al., 1999; Friesen & Kingstone,
The fact that other types of cues, like arrows, produce similar effects does not take away from the fact that people will orient their attention to where other people are looking. In other words, regardless of results with arrow cueing, the gaze cueing studies confirm that the eyes are important cues to the direction of another person’s attention.

A very different interpretation is that the gaze cueing paradigm is not capturing much about social attention. Rather, it measures more generally how spatial attention responds to any central directional cue. Because similar cueing effects are found for gaze, arrows, words (Hommel et al., 2001), and even wagging tongues (Downing et al., 2004), it may be that gaze cues are just one of many stimuli that vary on the critical dimension of direction. In this view, reflexive orienting for central gaze cues has little to do with social attention per se. That is, reflexive gaze cueing effects may not reflect the processes that support joint attention in the real world. What they do reflect is how attention shifts when a directional cue is presented at fixation (which may play only a small role in real world joint attention). This latter interpretation is supported by a recent study by Okada et al., (2002), who tested three autistic individuals who showed no joint attention behaviours in real world social situations. Despite their severe impairments with everyday joint attention, these individuals showed normal gaze cueing effects in the laboratory.

Perhaps more important than determining which of these two interpretations is ultimately correct is the appreciation that the gaze cuing paradigm is not capturing much of what makes eyes important social stimuli in everyday life. What is needed is a new approach for studying social attention. The present investigation represents an initial step toward achieving this aim.
Figure 7.1. The scenes used in the experiment. (A). Scenes with eyes and arrows, (B), Scenes with large arrows, (C) Scene without people.
Figure 7.2. Scenes overlaid with all observers’ fixations. There was a clear preference for the people in the scene, particularly their faces and eyes. Fewer fixations went to the arrows.
Figure 7.3. Fixation proportions for the scenes with eyes and arrows, as a function of region.
Figure 7.4. Fixation proportions for the scenes with large arrows, as a function of region.
Figure 7.5. Fixation proportions for the scenes with no people, as a function of region.
Figure 7.6. Proportion of first fixations in the scenes with eyes and arrows, as a function of region.
Figure 7.7. Proportion of first fixations in the scenes with large arrows, as a function of region.
Figure 7.8. Proportion of first fixations in the scene with no people, as a function of region.


Birmingham, E., Bischof, W.F., & Kingstone, A. (under review (a)). Is there a default bias to select the eyes? Under review at *Quarterly Journal of Experimental Psychology*.

Birmingham, E., Bischof, W.F., & Kingstone, A. (under review (b)). Saliency does not account for fixations to eyes within social scenes. Under review at *Vision Research*.


CHAPTER 8

General Discussion
Folk knowledge suggests that we are very interested in the attention of other people. This intuition that we care about the attentional states of others has led a flurry of research on social attention. While there are several cues to the direction of another person’s attention (e.g., gaze direction, head position, body position, pointing gestures), current models of social attention posit that eye gaze direction is the most important social attention cue (Baron-Cohen, 1994, 1995; Perrett et al., 1992).

Inherent in these models is the notion that social attention is composed of at least two distinct stages. One stage involves the selection of another person’s eye gaze as a key social stimulus, and the other stage involves the shifting of attention to where that person is looking. Indeed, it is often assumed that these two stages are obligatory in nature (e.g., Baron-Cohen, 1994). That is, humans have a fundamental tendency to preferentially select eyes from their environment and to shift attention to where those eyes are looking.

While the vast majority of research on social attention has focused on the latter of these two processes, i.e., the shift of attention to a gazed-at location, little research has been conducted on what social cues humans select from complex natural scenes. For instance, the popular gaze-cueing paradigm, which is used to study shifts of attention to gazed-at locations, is not designed to study the selection of gaze cues from the environment. This is because typically the gaze cue is presented in an otherwise empty visual field, effectively precluding any selection process on the part of the participant. And, while face-scanning studies certainly suggest that observers prefer to select the eyes of others (Althoff & Cohen, 1999; Henderson, Falk, Minut & Dyer, 2001; Walker-Smith, Gale & Findlay, 1977; Yarbus, 1967), these studies suffer from a similar limitation, in that observers are presented with a single face in an otherwise empty visual field. The finding that observers scan the eyes of isolated faces may be explained by the fact that in such an impoverished display the eyes are the most visually salient items on the screen (Itti & Koch, 2000). Figure 1.1 illustrates this possibility. Thus, finding that eyes attract fixations could be accounted for by a simple bottom-up saliency model of attentional control (Itti, Koch, & Niebur, 1998; Itti & Koch, 2000; Koch & Ullman, 1985) without any
reference to the notion that eyes are selected because of the social information that they convey, i.e., another person's attentional state, their intentions, emotions, and the like.

Almost no research has been conducted on the selection of eyes in more complex visual displays. One notable exception is the early work of Yarbus (1967) which found that when observers were shown images of a whole person, the preference to look at the eyes appeared to be largely diminished. However, when Yarbus presented a scene with multiple people, observers appeared to show a strong interest in the faces (and possibly the eyes, see Figure 1.2). Similarly, a study by Klin et al. (2002) has suggested that healthy observers prefer to look at the eyes when presented with complex scenes. However, as with Yarbus, there are several methodological concerns with Klin et al.'s study which obscure the extent to which observers were selecting the eyes relative to other regions in the scene (see Chapter 1 and Chapter 2 for a detailed description of these limitations of the Yarbus and Klin data).

The central aim of the present dissertation was to begin to examine in an empirically rigorous manner the issue of attentional selection of social stimuli, particularly the eyes, from complex natural scenes. Five main research questions were raised in Chapter 1. Below I consider the data that my investigations bring to each of these questions, the answers they suggest, and the directions that my studies suggest for future research.

1. Is there a preferential bias to select eyes from complex social scenes?

The present dissertation started out by asking whether eyes are preferentially selected from complex visual scenes. The answer is a most definitive "yes". Across the six studies of this dissertation, and for a variety of complex scenes and tasks, observers showed a preferential bias to selectively attend to the eyes of others.

A key finding of the present work is that the preference for eyes emerges early on, within the first one-to-two fixations (Chapters 3, 4 and 5), and most certainly within the first second of viewing (Chapter 2). These data could certainly serve as support for the idea of an innate "Eye Direction Detector" that rapidly detects eyes from the environment (Baron-Cohen, 1994).
Rapidly detecting the presence of eyes and using eye gaze as social signal may have been a necessary adaptation to the many changes happening throughout primate evolution (Emery, 2000). Given these changes, such as a flatter face and more prominent, higher contrast eyes, it would have been very adaptive to be able to detect where other individuals were looking based on their eye gaze. Consistent with this notion, many non-human primates and non-primate animals are also highly sensitive to the presence of eyes and eye-like stimuli and appear to be very sensitive to where those eyes are looking (Emery, 2000; Shepard & Platt, 2008).

On the other hand, our findings may simply reflect a behaviour that has been shaped by the strong reward value attached to quickly detecting and selecting eyes from the environment. For instance, babies learn that following an adult’s eye gaze often leads to interesting objects in the environment (Corkum & Moore, 1998), and seem to respond more positively when an adult makes eye contact than when an adult looks away. Indeed, babies prefer faces with direct gaze (Farroni et al., 2002) and smile more when a caregiver makes eye contact (Hains & Muir, 1996). Thus, the findings of the present work, showing that people orient to eyes rapidly and continue to select them over time, may reflect this social reward learning.

Based on past nature-nurture discussions, it is most likely that the answer is not to be found at either extreme, but between the two positions. Specifically, it seems reasonable that humans have developed a brain architecture that is finely-tuned for face and eye processing, but that the healthy development of this social attention network depends on learning what social cues like eyes convey and represent as a child grows and develops (Corkum & Moore, 1998).

2. What factors influence this selection process?

Two key factors were found to modulate the extent that eyes were selected: the social content and activity of the scenes presented, and the task being performed by the observer. Chapter 2 showed that the selection of eyes was particularly enhanced in scenes in which examining gaze might be especially useful, that is, in scenes with high social content (multiple
people) and activity. This suggested that gaze information is critical to understanding the
meaning of actions within social situations. Chapter 3 replicated the effects of social content and
activity found in Chapter 2, and additionally showed that observers who were asked describe
where people in the scenes were directing their attention enhanced their fixations to the eyes
relative to two control groups (observers asked to ‘describe the scenes’ or ‘look at the scenes’).
Chapter 4 also found that tasks requiring observers to process social aspects of the scene, such
as where people were attending, or what they were thinking or feeling, enhanced fixations to the
eyes relative to non-social tasks (studying the meaning, informative aspects, or uninformative
aspects of the scenes). When considered collectively these results support the general notion
that eyes are understood to provide important socially communicative information. Moreover,
these findings dovetail with previous studies showing that observers are able to accurately judge
complex mental states from eye information alone (Baron-Cohen et al., 1997).

3. To what extent does the preferential bias to select the eyes generalize to different tasks and
situations?

Chapter 4 asked whether there really is a fundamental bias to select the eyes, or
whether previous task instructions (e.g., those in Chapter 3) may have inadvertently encouraged
observers to look at the people in the scenes. Thus, different groups of observers were given
different task instructions, some that were clearly about the people in the scene (e.g., their
emotional states) and some that were more general in nature (e.g., the informative items in the
scenes). The results showed that observers in all task conditions, even observers who were
asked to study the uninformative aspects of the scenes, showed a strong bias to select the
eyes. This bias began early (within the first fixation) and persisted across the viewing session. It
was only later (by 8 seconds) that task had a significant effect on performance, with social tasks
leading to increased fixations to the eyes relative to more general tasks. When taken together,
these findings suggested that the fundamental bias to select eyes reflects observers’ default
interest in eyes as important social-communicative stimuli.
Chapter 5 asked whether observers perceive the eyes to be important both for encoding and remembering scenes. That is, until the study reported in Chapter 5, all the investigations had examined performance during observers' initial viewing of the scenes, i.e., when the scenes were being encoded. It was unclear whether this basic interest in eyes would persist during scene recognition. The results showed that observers who were told initially that their memory for the scenes would be tested later (Told group) looked more at the eyes of the people in the scenes than those that were not aware of the fact that there would later be a memory test (Not Told group). Most importantly, it was found that when recognizing the scenes, observers tended to focus on the same scene regions that they fixated when they first viewed the scenes, with observers in the Told group again fixating the eyes more often than the Not Told group during the memory test. This suggests that interest in the eyes during scene encoding, including incidental scene encoding (Not Told group), carry over to scene recognition. Thus again the data suggest that while there is a broad and fundamental interest in the eyes of others. This bias can be modulated by top-down strategies that are remarkably stable and appear to be established at the time of encoding.

4. What is the role of visual saliency in driving fixations to the eyes within complex social scenes?

Chapter 6 assessed the role of saliency in driving observers to fixate the eyes in complex scenes with people. Saliency maps were computed for the scenes from four experiments (Chapters 2-5) by a standard computational model of saliency (Itti & Koch, 2000). The main prediction of this model is that salient items should be fixated more often than non-salient items, particularly early on in viewing (e.g., the first fixation). In contrast to this prediction, the data showed that saliency provided an incredibly poor account of the data. Across all experiments, the saliency values for the first-fixated locations were extremely low (i.e., not salient according to the model), and no greater than what would be expected by chance alone. In addition, it was found that the saliency values for the eye regions in the
scenes were low, and no more salient than any other region. Thus, visual saliency does not account for observers’ strong bias to select the eyes in the scenes, nor does it account for fixation behaviour in general.

5. How does studying the selection of social cues shed light on past controversies in the social attention literature?

If gaze is such a unique and important cue to social attention, then why do other non-biological cues like arrows also produce reflexive orienting effects? Gaze and arrow cueing studies of spatial orienting have shown that eyes and arrows produce virtually identical effects on shifts of spatial attention (Hommel, Pratt, Colzato, & Godijn, 2001; Pratt & Hommel, 2003; Ristic, Friesen, & Kingstone, 2002; Ristic, Wright & Kingstone, 2007; Tipples, 2002). This has led some researchers to suggest that arrows are highly socially important stimuli that may be comparable to eyes (e.g., Kingstone et al., 2003). Certainly, arrows are used frequently by humans to direct attention, and we see them all the time in our everyday life – on road signs, on the pavement, as turning signals at traffic intersections, etc. Thus, one interpretation of the results from cueing studies is that eyes and arrows are equally socially relevant, and thus given equal priority by the attention system. However, this interpretation does not fit with the general position in the social attention literature that eyes may be unique social stimuli. As a result there has been considerable debate about the similarity between gaze and arrow cues, both at a behavioural level (Friesen et al., 2004; Tipples, 2008), and at a neural level (Kingstone et al., 2000; Hietanen et al., 2006; Ristic et al., 2002; Tipper, Giesbrecht, Handy & Kingstone, 2008; Vuilleumier, 2008).

To shed light on this controversy, I reasoned that measures of selection might be more sensitive to differences in these cues (e.g., the unique social importance of eyes), and as such they might reveal an attentional (selection) priority for eyes over arrows.

Thus, Chapter 7 showed observers real world scenes containing people and arrows, and measured what people preferred to look at. The results were unequivocal: eyes were fixated far
more often than arrows, which were rarely fixated. Indeed, arrows did not receive much
attention at all, and varying the size of the arrows or removing people from the scenes
altogether did not enhance arrow selection. Additionally, whereas the eyes were often fixated
first (along with heads and text), arrows were never fixated first. Thus, arrows and eyes were
not given equal priority by the attention system in terms of selection.

**Implications**

There are several important implications of the present work. First, the methodology that
was developed here for measuring and quantifying the preferential bias for social stimuli has
produced a novel and robust tool for examining social attention. In particular, the selection
component of social attention, which has received little empirical consideration in the literature,
can be measured readily by eye monitoring observers while they view social scenes. Using this
approach, I uncovered a robust, fundamental preference for eyes that persists across various
tasks. This bias, however, can be enhanced by tasks that instruct observers to process social
aspects of the scenes (e.g., attentional states), or by presenting scenes that contain high levels
of social content and activity. Collectively these data suggest that observers have a preferential
bias to look at the eyes of others because they provide important social information. While the
idea of measuring task-effects on fixation behaviour dates back to the work of Yarbus (1967),
the studies of the present dissertation represent the first systematic and rigorous application of
task and scene content manipulation to tap into the mechanisms of social attention.

One of the main goals of using this approach was to examine observers' selection of
social stimuli -- specifically the eyes -- without introducing the fundamental flaw of having the
eyes as the most salient items in the display. As described before, it is often very difficult to
assess what can be taken away from face scanning studies because the faces in these studies
are typically isolated from their normal context, creating an impoverished situation in which the
eyes are potentially the most visually salient part of the display (Althoff & Cohen, 1999;
Henderson, Falk, Minut & Dyer, 2001; Walker-Smith, Gale & Findlay, 1977). Indeed, it was
precisely for this reason that the analyses reported in Chapter 6 were conducted, i.e., to
determine what role, if any, visual saliency plays in the selection bias for eyes. Finding that visual saliency does not account for fixations to the eyes in my studies strengthens the conclusion that observers in the present studies were interested in the eyes because of the social information that they afforded to the observer.

Do the present data mean that the preference for eyes that has been measured in the isolated face scanning studies reflects social attention? In other words, is one justified to conclude that because the present dissertation has found that observers do look to the eyes for social information, that this means that they also look to the eyes for social information when presented with an isolated face? While such a conclusion is tempting, there are several reasons that it must be viewed with caution.

One reason is that the selection process, which is the cornerstone to the present studies, is experimentally constrained in the face scanning studies that present observers with a face in isolation. Because the face is the only figure on the screen, observers are limited to selecting among the few items within the face (e.g., eyes, nose, mouth, chin). Thus, the more complex processes of selection that are measured in the present studies are not being tapped by the typical face scanning studies. A second reason is that by isolating the face from its surrounding environment, researchers are introducing the confound of feature saliency. That is, it may be that observers fixate the eyes because the eyes have the highest saliency value in the display and not because the eyes are conveying important social information. A third reason, related to the second, is that by isolating the face from its surroundings, one has not only introduced the confound of visual saliency into the display, but one may have removed much of the social value of the face. In other words, by embedding a face stimulus in an otherwise empty scene, one has also removed the face from its normal context and made it socially impoverished. In the real world, we rarely (if ever) interact with faces detached from their bodies – we generally interact with full-bodied people. As such, it is highly unlikely that participants in face scanning studies treat the face in front of them as a social agent, i.e., a real person with intentions and complex mental states. Of course, the ideal stimulus for an ecologically-valid social attention
experiment would certainly be a real living person engaged in social interaction. The social scene viewing approach used in the present work represents a first significant step toward research in this domain.

Another implication of the present dissertation is that it demonstrates the richness of measures of selection for studying social attention. The utility of measuring selection was demonstrated in Chapter 7, which discovered that while gaze and arrows may produce equivalent effects on shifts of attention in the context of the cueing paradigm (e.g., Ristic et al., 2002), observers do not select eyes and arrows to the same extent when they are embedded within more complex scenes – that is, observers select eyes over arrows.

It is perhaps worth noting that the idea that one can gain new insights into human behaviour by increasing display complexity, and the range of response options to a participant, is consistent with a new research approach called Cognitive Ethology (e.g., Kingstone, Smilek, Birmingham, Cameron & Bischof, 2005). This approach proposes that by freeing the individual to express his or her natural behaviour and interests, one can derive new insights into human cognition. To take a specific example, the gaze cuing paradigm has been unable to determine whether eyes and arrows are, or are not, equally important social cues. In fact, the evidence has suggested that they are equivalent. However, in these studies the stimulus is either a face or an arrow (they are never presented together, so there is no selection process), the stimulus is presented at fixation (so there is no eye movement required, in fact subjects are told not to move their eyes), and the response is to press a single key (there is only one available to the participant). The task is constrained in such a way because the experimenter wants to control all factors, except the one of interest -- eyes vs. arrows -- in order to maximize analytical power. In fact, a recent study eliminated all differences between arrows and eyes, including any difference in the stimulus itself, and found that an ambiguous stimulus, whether it is perceived as an eye or an arrow, engages the same brain networks (Tipper et al., 2008). In other words, the 'classic' research approach, of reducing the complexity of the situation, had led investigators to the conclusion that eyes and arrows are equivalent, both behaviourally and in the brain.
mechanisms they engage. In contrast, the principle of cognitive ethology suggests that new insights can be gained by increasing the richness of the display and the response options to the participant. In keeping with this approach, the study reported in Chapter 7 revealed that when eyes and arrows are embedded in complex natural scenes and an observer is permitted to fixate whatever s/he wants to look at, eyes and arrows are treated incredibly differently, with eyes being given a selection priority that greatly exceeds the selection priority granted to arrows.

**Future directions**

In keeping with the principle of cognitive ethology, future research investigations into social attention are likely to benefit by continuing to increase the stimulus complexity of experimental environments. For instance, a natural next step into the current research program would be to examine how people view natural scenes when the people within the scenes are in motion. Indeed, the research by Klin et al. (2002) certainly suggests that observers show an interest in the eyes when viewing movie clips. However, in addition to the concerns with the hand-coding scheme used by Klin et al. (2002), there are some issues with the content of the clips that beg future investigation. For instance, the authors deliberately chose clips that were “depleted of nonessential objects and events that might distract a viewer’s attention from the social action” (p. 811). Given that this resulted in close-up shots of highly charged social interactions, one might wonder whether the preference for eyes in their study was a product of either the faces of the people being isolated (echoing the problems that were raised with the face-scanning studies at the beginning of this thesis) and/or the specific situation being faced by the observer. Thus, a future study would be to present observers with full-body dynamic social situations that are more natural -- involving real people (not actors) -- in their everyday environments, and assess whether there is a bias to look at the eyes. Note that it would be critical to rule out the possibility that fixations might be attracted to the people (and eyes) in dynamic scenes simply because they are in motion (another potential problem with Kiln et al.’s study). Thus, instead of excluding ‘nonessential objects and events’ from the clips, I would deliberately include other objects and manipulate whether or not they are moving.
The increase in complexity from static to dynamic scenes would also afford the opportunity to naturally introduce other factors heretofore untested, e.g., the role of emotions of people in the scenes, the relative status of the people in the scenes, the presence of sound, speech, and so forth. Indeed, one might also be interested in knowing whether factors related to the observer, such as his/her cultural background, gender, or age, might alter viewing patterns within social scenes. For instance, there is accruing evidence that “Westerners” perceive people to be autonomous individuals with their own distinct feelings, goals, and attributes, whereas East Asians perceive people to be inseparable from other people, i.e., the group (Markus & Kitayama, 1991; Markus, Kitayama, & Heiman, 1996; Markus, Mullay, & Kitayama, 1997). Consistent with this, in a recent study Masuda et al., (2008) showed that when judging the emotion of a central cartoon character, Japanese participants were influenced by the emotions of other characters surrounding the central character, whereas Western participants were not. In addition, the authors found that this cultural difference in emotion judgment performance was reflected in differences in eye movements, with Japanese participants looking more often than Western participants at the surrounding characters in the scene. Given these findings, one potential future investigation would be to look for cultural differences in viewing preferences while observers judge the mental states of people in natural, real-world, social scenes. For instance, one might predict that cultural viewing pattern differences would be more pronounced when there are multiple people in the scene than when there is only one person in the scene. In addition, it would be interesting to know whether there are cultural differences in the specific preference for eyes, as opposed to other facial features, as useful conveyors of social information.

Another step is to move from eye monitoring people while they view images of people in motion to eye monitoring people while they view real people in motion. It is worth noting that to date, virtually all of the research in social attention (including the present dissertation) has been limited to situations involving images of people. By definition, these images of people cannot attend to the observer while the observer is attending to them. This stands in sharp contrast to
many situations in real life. Interestingly, while one might be tempted to predict that observers would look even more often at the eyes in real social situations than in images of social situations, the opposite could just as easily be true. For instance, while eye contact is a functional part of everyday social interactions, social norms indicate that it is rude to make excessive eye contact or to spend too much time looking at another person. Indeed, in some situations (e.g., being approached by a hostile person) it is clear that eye contact should be avoided altogether. This may be a key new factor for consideration in future studies of social attention that has not yet been touched on at all.

Finally, a new line of investigation would be to bring together the two key processes of selection and orienting. Past studies have focused on the shift of attention to gazed-at locations, to the exclusion of selection of gaze. The present dissertation has focused on selection of gaze, largely to the exclusion of orienting to gazed-at locations. Bringing these two processes together, for instance, examining the selection of social stimuli and how this selection process affects where attention is directed next, is an exciting future program of research.
References


APPENDIX I

UBC Behavioural Research Ethics Board Certificate of Approval
The University of British Columbia  
Office of Research Services  
Behavioural Research Ethics Board  
Suite 102, 6190 Agronomy Road, Vancouver, B.C. V6T 1Z3

CERTIFICATE OF APPROVAL - MINIMAL RISK AMENDMENT

<table>
<thead>
<tr>
<th>PRINCIPAL INVESTIGATOR:</th>
<th>DEPARTMENT:</th>
<th>UBC BREB NUMBER:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alan Kingstone</td>
<td>UBC/Arts/Psychology, Department of</td>
<td>H99-80321</td>
</tr>
</tbody>
</table>

INSTITUTION(S) WHERE RESEARCH WILL BE CARRIED OUT:

<table>
<thead>
<tr>
<th>Institution</th>
<th>Site</th>
</tr>
</thead>
<tbody>
<tr>
<td>UBC</td>
<td>Vancouver (excludes UBC Hospital)</td>
</tr>
<tr>
<td>Other locations where the research will be conducted:</td>
<td>N/A</td>
</tr>
</tbody>
</table>

CO-INVESTIGATOR(S):

Elina Birmingham  
Kirsten Dalrymple  
Wieske Van Zoest  
Joseph Chisholm

SPONSORING AGENCIES:

Natural Sciences and Engineering Research Council of Canada (NSERC) - "Components of Human Selective Attention" - "Toward an Integrative Science"  
Social Sciences and Humanities Research Council of Canada (SSHRC) - "Orsil title - Inferring Attention in Social Situations: A Cognitive Ethology"

PROJECT TITLE:

Cognitive Ethology: 1-20

Expiry Date - Approval of an amendment does not change the expiry date on the current UBC BREB approval of this study. An application for renewal is required on or before: July 30, 2008

AMENDMENT(S):

<table>
<thead>
<tr>
<th>AMENDMENT APPROVAL DATE:</th>
<th>February 26, 2008</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Document Name</th>
<th>Version</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consent Forms:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>consent</td>
<td>2</td>
<td>February 13, 2008</td>
</tr>
</tbody>
</table>

The amendment(s) and the document(s) listed above have been reviewed and the procedures were found to be acceptable on ethical grounds for research involving human subjects.
Approval is issued on behalf of the Behavioural Research Ethics Board and signed electronically by one of the following:

Dr. M. Judith Lynam, Chair
Dr. Ken Craig, Chair
Dr. Jim Rupert, Associate Chair
Dr. Laurie Ford, Associate Chair
Dr. Daniel Salhani, Associate Chair
Dr. Anita Ho, Associate Chair