ROYAL HOLLOWAY UNIVERSITY OF LONDON

DOCTORAL THESIS

---

# The Potential for Unknowingly Disclosing Personal Information via Eye Tracking Technology

---

*Author:*
Callum WOODS

*Supervisors:*
Dr. Szonya DURANT,
Prof. Zhiyuan LUO,
Prof. Dawn WATLING

*A thesis submitted in fulfillment of the requirements for the degree of Doctor of Philosophy*

*in the*

Department of Psychology

# Declaration of Authorship

I, Callum WOODS, declare that this thesis titled, The Potential for Unknowingly Disclosing Personal Information via Eye Tracking Technology and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- No part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

Date:

ROYAL HOLLOWAY UNIVERSITY OF LONDON

# *Abstract*

Department of Psychology

Doctor of Philosophy

**The Potential for Unknowingly Disclosing Personal Information via Eye Tracking Technology**

by Callum WOODS

We evaluate whether personal information, such as an individual's personality, gender, or self-esteem can be predicted from their visual behaviour upon social networking site (SNS) based content. This SNS context provides an ecologically valid, and novel, visual environment and behaviour upon such sites has been found to reflect a wide range of personal attributes. Our novel contribution to the literature is to highlight that, through the use of machine learning techniques, visual behaviour provides insight into a range of personality traits and personal attributes within very short (sub minute) time scales. This is in contrast to previous approaches that aggregate digital logs of SNS behaviour across weeks, months or years of use to make similar predictions. Furthermore, we evaluate which types of visual behaviour are most informative when predicting personal information and find that, in certain situations, it appears that it is not critical to know the type of content being displayed upon the page. We highlight that this has important implications for privacy, especially with eye tracking becoming increasingly popular as a way for users to interact with their computer.

# *Acknowledgements*

This thesis represents the end product of a three year journey. It has been life changing, and I could not possibly express my gratitude for the opportunities I have had. The path (whilst not always paved!) has led to many amazing experiences that have been key to the development of this thesis and myself as an individual. This would not have possible without my supervision team (Szonya Durant, Zhiyuan Lou and Dawn Watling), who I would like to thank for being constantly supportive and dealing with my frequent bewilderment with grace and patience. A special thank you goes out to my lead supervisor Szonya Durant for her dedication to speedy reviews and always having time to discuss any problems big or small. It is a big job taking on a PhD student, and I realised this more as time went on! I would also like to thank the members of the PhD cohort for providing constant support (I'm looking at you, Dave and Craig). Finally, this process has brought a rather wonderful Sahira van de Wouw into my life, who I would like to thank for keeping me sane in the harder times, and for bringing magic to the best of times.

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Incorporating eye movements into human computer interactions may lead to users unknowingly disclosing personal information

In the following chapter, I review relevant literature upon how the physical properties of the stimulus, task, and previous experience can influence how people view a scene. This provides a framework by which to lend support to the fundamental proposition that the thesis rests upon - that an individual's oculomotor behaviour may provide a reflection of aspects of their personality and personal attributes. Building upon this literature, I then review the current state of the art in predicting personal attributes from visual behaviour to illustrate that the incorporation of eye movements into everyday human-computer interactions may lead to the user disclosing more about themselves than they realise. Finally, to understand the potential risk to privacy posed by the prediction of psychological traits and illustrate the real-world relevance of this thesis, I will briefly outline how personal information may be used to manipulate human behaviour and why this is valued by interested parties.

To understand the type of information and inferences that can be drawn from visual behaviour, one must first understand the mechanisms driving the expression of visual behaviour. As such, I start by briefly outlining visual attention and the factors which influence it. This progresses to describe the link between reward and visual behaviour, explaining how the user's preferences can influence the distribution of their visual attention. Finally, to understand the potential risk to privacy posed by the prediction of psychological traits, I discuss the potential social, societal, and privacy issues surrounding the ability to implicitly predict an individual's personality traits from oculomotor behaviour. This highlights the relevance of the thesis topic and includes examples where psychological profiling has been used to direct public opinion and influence individuals into making purchasing decisions. A particular issue highlighted is that the temporal granularity of oculomotor data may provide novel privacy issues (and solutions) in comparison to more traditional psychological profiling methods. This leads to three main topics: eye movements and attention, the ability to predict personal attributes from visual behaviour, and the potential risks emerging from implicit disclosure of personal information via eye tracking technology.

## 1.1 Eye Movements and Attention

### 1.1.1 Oculomotor Events: Fixations, Saccades and Blinks

Visual acuity refers to the smallest detail which can be seen under given conditions (Pirenne, 1962) and is greatest (e.g. we can see smaller items more clearly) at a central region of the eye called the fovea, which spans roughly two degrees of visual angle (Pirenne, 1962). This means that when we direct our eyes at a location (e.g. light from the location lands on the fovea), that region is resolved with the greatest visual acuity, with visual acuity reducing eccentrically from this point. This occurs as the physiological composition of the eye (and correspondingly, the functional contribution to vision) changes with eccentricity from the fovea (Battista, Kalloniatis, & Metha, 2005; Pirenne, 1962). To elaborate, the fovea excels at high quality inspection within bright conditions as it is densely populated by photoreceptors, in particular cone cells (a type of photoreceptor that provide colour vision and works best in bright lighting conditions; Pirenne, 1962). The fovea receives preferential processing, with the central $2.5^o$ of visual angle being represented by 25% of the visual cortex (De Valois & De Valois, 1980). The counterpart to cone cells are rod cells, which excel at detecting motion and work best within dim lighting conditions. These rod cells are most prolific at more eccentric visual angles, and lead to the eye being sensitive to dim light and movement within the periphery (Carrasco, 2011).

The eyes themselves are part of the central nervous system and employ a variety of filtering mechanisms, effectively processing the photon-level data before it is transmitted to the cortex via ganglion cells (Gollisch & Meister, 2010). This processing occurs via roughly twenty different ganglion cell populations, which form mosaic structures across the eye to filter the visual scene (in parallel) before passing this information to the cortex (Robles, Laurell, & Baier, 2014). Here, further processing occurs to extract features such as colour, contrast, luminance, movement and orientation (De Valois & De Valois, 1980; Itti & Koch, 2001; Tatler, Baddeley, & Gilchrist, 2005; Treue, 2003) among others. Together, these features define the physical saliency of objects or locations within the visual scene - a topic explored in depth later in this introduction. Due to the highly parallel nature of visual processing, the visual scene is represented along several spatial scales, from coarse to fine grained (De Valois & De Valois, 1980) and also over time (Cohen & Newsome, 2009).

It has been suggested that the most important function of the visual system is to rapidly foveate (i.e. locate in the fovea) salient objects, enabling the organism to efficiently interact with the environment (Itti & Koch, 2001), which is thought to convey an evolutionary advantage in detecting prey, mates, or predators (Emery, 2000). As the visual system requires an image to be held relatively stable on the retina for best perception, this creates a diphasic cycle as the eye balances the need to stabilise the visual scene with locating new regions of the visual scene within the highest region of visual acuity. This produces two distinct behaviours, when the eye is held stable (fixation) and when it is moving to a new location (saccade). The average human makes roughly 2-3 fixations a second, which usually last around 300ms (Kowler, 2011). Saccades are much briefer, often taking only 30-60ms and reaching velocities up to 500°/s (Holmqvist et al., 2011). The visual system is mostly unable to process the scene during these events. For this reason, the time and location of fixations provide information as to when and where visual information is available for cognitive processing; albeit, after saccadic movement, the eye wobbles for a time (post-saccadic oscillation), with visual detection reduced before (50% at -20ms) and after (50% at +75ms) saccade onset (Kowler, 2011; Volkmann, Schick, & Riggs, 1968).

Together, this literature demonstrates that the visual scene is represented by extracted features upon multiple spatial scales (Gollisch & Meister, 2010; Robles et al., 2014). This leads to being able to conceptualise visual input to the brain as being a series of snapshots, which are later stitched together to form the coherent stream of visual information that we consciously perceive (Hayhoe & Ballard, 2005; Holmqvist et al., 2011; Kredel, Vater, Klostermann, & Hossner, 2017). We have identified that fixations indicate when the visual scene is available for cognitive processing, and that saccades indicate the journey to locating a new visual location within the region of highest visual acuity. As such, the focus of this thesis is upon these two events, as they can be used to represent how the individual explores the visual environment. For completeness, we note that a special case is during smooth pursuit, where the eye is fixating upon a moving object. To keep the image stable on the fovea, the eye makes only small 'smooth' adjustments, maintaining the visual scene for cognitive processing (Holmqvist et al., 2011). The impact of smooth pursuit upon the algorithmic detection of fixation and saccadic events in the context of web page style stimulus is addressed within the methodology section. In this current section, I have also identified additional small oculomotor events that, as they are co-incidental to the exploration of the visual scene, are outside of the scope of this thesis (such as post-saccadic oscillations). Finally, the eye must be kept moist and clean, and the eyelid frequently coats the cornea to preserve vision. This leads to blinks, during which the visual scene is obscured by the eyelid and is not available for processing.

### 1.1.2 Bottom-up versus top-down influences on visual attention

The eye receives an incredible throughput of information. However, we only have limited resources for interpreting this data. This leads to the discovery that we may fail to consciously perceive aspects clearly detectable within the visual scene (Kowler, 2011; Serences & Yantis, 2006; Simons & Chabris, 1999; Treue, 2003). That we consciously perceive only a small fraction of the visual scene before us is perhaps most famously illustrated by the inattentional blindness experiments of Simons and Chabris (1999). By having a person in a hairy black gorilla suit casually stroll through the attended visual scene, the authors were able to demonstrate that visual objects, even when presented in clear view, may not reach conscious perception without attention. The gorilla was only reliably noticed when the demands of the experimental task were congruent (e.g. tracking black objects) with the cognitive task. If not all information that reaches the eyes is perceived and processed, a mechanism must act as a gateway to conscious perception. This mechanism is referred to as visual attention, defined here as the process of selecting a subset of stimuli out of all available visual stimuli (Itti & Koch, 2001). Within this section, I will introduce factors which influence how individuals select which stimuli to attend to, thus providing the theoretical basis for how individual differences can influence patterns of eye movements across a scene.

Saliency refers to how important or noticeable an object is within the visual scene. In the vision literature, a distinction is often made between the physical aspects of the stimuli that influence its saliency (low-level), and cognitive factors, such as task requirements (top-down influences). Low-level saliency refers to the relative importance, or notability, of an object within a scene due to its physical features. For example, a red circle will 'pop out' and be very noticeable within a group of green circles, due to its physical property of having a high relative colour contrast (Maunsell & Treue, 2006; Treue, 2003). This is a stable property and inherent within the composition of the visual scene. Common physical attributes that may influence the likelihood of an object receiving visual attention include its colour, luminance, contrast, movement, and orientation (Itti & Koch, 2001).

In contrast, an excellent example of a top-down influence upon visual attention is present within the gorilla example given above, as the participants were much more likely to detect the presence of the gorilla when they were asked to track a black object through the visual scene (Simons & Chabris, 1999). As such, without influencing any of the physical attributes of the visual scene, an external factor such as the task (e.g., tracking a black object) may alter the distribution of visual attention within the scene. That the cognitive task of the participant has a profound influence upon which elements receive visual attention (as measured by eye gaze positions) in a scene was first identified within the experiments of Yarbus (Tatler, Wade, Kwan, Findlay, & Velichkovsky, 2010), where, in a simple paradigm, participants were given different cognitive tasks to perform whilst viewing a painting. For example, to identify if the family was wealthy. By tracking the subsequent eye movements, it was discovered that, dependent upon the task, the subject would employ a very different series of fixations (also termed scan paths) for different objectives (Tatler et al., 2010). The influence of cognitive requirements upon visual saliency are often referred to as 'high level', or 'top-down' influences as they are thought to originate within cognitive regions that are later on in the processing stream (Funahashi, 2006; McGinty, Rangel, & Newsome, 2016).

Itti and Koch (2001) proposed an influential computational model that views saliency as being represented through the explicit cortical representation of physically derived features; a single 'saliency map' informed by the output of both retinal and cortical processing upon multiple scales from which the winner takes all (e.g. the winner receives visual attention). To elaborate, fundamental to this theory is the concept of spatiotopic mapping, where the relative spatial locations of objects are preserved within the visual cortical areas (De Valois & De Valois, 1980; McGinty et al., 2016; Robles et al., 2014; Tatler et al., 2005; Tatler, Hayhoe, Land, & Ballard, 2011). Thus, the cortex is able to represent a spatially preserved scene, derive the region of highest salience, trigger a saccade (via the superior colliculus; Robinson, 1972) and then inhibit the return of the eye to this location (De Valois & De Valois, 1980; Itti & Koch, 2001; Volkmann et al., 1968) to ensure fixations capture novel information. In this model, downstream cortical regions (e.g. hierarchically superior) produce goal-related eye movement patterns through biasing the relative weight (and thus priority) of specific types of input. This acts to suppress fixations upon task-irrelevant regions, whilst simultaneously enhancing the salience of regions likely to be task-relevant. This creates a dynamic closed feedback loop whereby objects with little physical (low level) saliency can effectively compete for attention with physically salient objects when the former are task-relevant.

In summary, top-down control has a large influence upon how an individual visually interacts with the scene in front of them (Tatler et al., 2010). This dynamic goal directed allocation of attention acts as a mechanism to enhance the cortical representation of task relevant objects in the visual scene (Itti & Koch, 2001; Serences & Yantis, 2006), thus the task can influence how an individual allocates their visual attention across the visual scene (e.g., as seen in: Simons & Chabris, 1999). It is this top-down re-weighting mechanism which provides the theoretical justification for why an individual's personal attributes may be reflected in the distribution of their visual attention, which is explored further in the next section.

### 1.1.3 Visual routines: Learning where to look

Hayhoe and Ballard (2005) outline three main developments which have contributed substantially to our understanding of the role of eye movements in reflecting cognitive function. Mentioned at the start of this segment, the first relates to the findings of Yarbus and subsequent researchers in identifying the influence of the task upon visual behaviour (Tatler et al., 2010). The second regards the role of reward in learning and guiding visual behaviour, and the third relates to developments in the field of reinforcement learning. In this section, we expand upon these three key themes, explaining how each provides a mechanism by which visual behaviour can be influenced by personal factors, whilst highlighting relevant current research.

That individuals prioritise fixating the location that offers the best support for task-relevant actions (task-congruent locations), rather than necessarily locations of high physical (e.g. low-level) salience, has been referred to as a 'just in time' strategy, and has been demonstrated to occur not just when viewing screens, but also both within virtual reality (Bhorkar, 2017; Milgram & Drascic, 2001) and natural environments (Hayhoe & Ballard, 2005; Kredel et al., 2017). In short, we fixate on what is relevant from moment-to-moment as the task progresses, and move to the next location when we have obtained sufficient visual information to proceed. This also includes predictive fixations to where we believe visual information will be. This is illustrated by Land and McLeod (2000) who tracked the eye movements of batsmen playing cricket. They found that both amateur and professional players fixate upon where they believe the bowled ball will land, rather than trying to continuously track the fast moving object. However, the analysis also revealed that professional players reached this point earlier in the kinematic sequence than non-professionals. It was also suggested that the professional players relied on extracting advance cues from much earlier in the bowler's sequence than amateurs (Land & McLeod, 2000). This shows that professional batsmen may have a unique oculomotor signature (whilst playing) that is distinguishable from amateurs, and raises a pertinent question: How do these professionals know where to look, and when to do so? It becomes clear that a learning process is at play here, and that there is plasticity to learn new strategies and patterns of viewing based upon prior experience. The ability to select and deploy different sets of task-specific eye movements, often referred to as visual routines (De Valois & De Valois, 1980; Hayhoe & Ballard, 2005; Summerfield & Egner, 2009; Vuilleumier, Henson, Driver, Dolan, & Orban, 2002), suggests that differences in experience and learning play a part in inducing individual variance in viewing patterns. This leads to the next question, how do we learn the visual routine and when to deploy it?

### 1.1.4 Reward, Learning and Gaze Bias

The second development Hayhoe and Ballard (2005) identified relates to the role of reward in motivating and guiding eye movements, with Anderson (2013) noting that reward related stimuli will not always be physically salient, nor expected (e.g. task-relevant). This implies that, in order for an organism to maximise reward and minimise loss, a separate mechanism is needed to allow the organism to learn to direct attention towards such stimuli even when they are non-salient and irrelevant to the current task. There is accumulating evidence for viewing reward as a separate and distinct mechanism, which biases attention as a form of perceptual learning. For example, given a repetitive perceptual discrimination task, participants will become more sensitive to the manipulation and faster to react as trials progress (Hamamé, Cosmelli, Henriquez, & Aboitiz, 2011). This is thought to occur due to a linear combination of increased connectivity between sensory and decision making regions, and early visual areas becoming attuned (more sensitive to) to the relevant stimulus properties (Chen et al., 2015). As such, reward acts as a mechanism to increase the efficiency of selection in visual search (Chen et al., 2015; Niu, Todd, & Anderson, 2012; Schotter, Berry, McKenzie, & Rayner, 2010) by priming the selection of previously rewarding stimuli in that environment. This leads to attentional capture being driven in part by reward history, learnt from observing the outcomes of previous encounters (Chelazzi, Perlato, Santandrea, & Della Libera, 2013; Kelley & Yantis, 2010). This supports the suggestion that reward mediated attentional capture allows rewarding stimuli to effectively compete with surrounding objects that may be both visually salient and task relevant, an aspect which has been understated in more traditional models of attention (Anderson, 2013).

The neural correlate of reward is encoded within the firing pattern of dopaminergic neurons, such as those within the substantia nigra (Alexander, DeLong, & Strick, 1986; Joshua, Adler, Mitelman, Vaadia, & Bergman, 2008). Of note, is that reward is encoded as the difference between the predicted and actual outcome. For example, an unexpected reward will elicit a much stronger signal than an expected reward. Similarly, the lack of an expected reward will be encoded in the response of dopamine neurons in the substantia nigra and the ventral tegmental area (Ikemoto, Yang, & Tan, 2015). This can be thought of as a bidirectional prediction error signal (Chang et al., 2015) and is thought to play a role in mediating the expression of behaviours that seek to optimise reward (Montague, Hyman, & Cohen, 2004). Stimulation of the substantia nigra reduces firing rates in the superior colliculus (Liu & Basso, 2008), a region directly responsible for initiating eye movements (Robinson, 1972). Direct projections have been traced from the substantia nigra to the superior colliculus (Mai, Majtanik, & Paxinos, 2015) and there is also an argument that, as part of the basal ganglia, dopaminergic neurons in the substantia nigra also influence reward, mood and approach motivation (Ikemoto et al., 2015).

Together, the above literature indicates how the structures that encode the prediction and outcome of reward (and aversion, see: Ilango et al., 2014) directly impact, and play a fundamental role in, the initiation of eye movement. The regions mentioned (e.g., superior colliculus) can be considered low-level (bottom-up), but the classification of a stimulus/outcome as rewarding involves associative (top-down) input (Ramnani & Owen, 2004). This creates ambiguity in whether reward is a top-down or bottom-up moderator of attention. When the association between a stimulus and reward is not yet established, this can be conceptualised as a top-down process whereby the semantic evaluation of the outcome is evaluated as positive in the context. However, once established, this persistent stimulus-response association can be considered part of a non-volitional bottom-up process. The result is that, due to being directly influenced by reward-related neural mechanisms, each individual develops a unique profile of stimulus-response associations, learnt from experience, that leads to them subconsciously expressing eye movements that reflect aspects of what an individual finds rewarding or adverse (unfavourable) in the visual environment. In short, eye movement is motivated by, and may reflect, the individual's preferences and cognitive biases (Armstrong & Olatunji, 2012; Saito, Nouchi, Kinjo, & Kawashima, 2017).

This relates the work of Campos, Frankel, and Camras (2004), who proposed that few or no environmental cues intrinsically evoke emotion. Instead, an emotional response is entwined with the individual's construed significance, linking events to a personal value or relevance. When describing the interaction between emotion and emotional regulation, Campos et al. (2004) identify that an individual's choice of environment influences their emotions; an action they term nichepicking. Nichepicking is described broadly, such as an individual choosing to spend their time either at a theme park or a library. Based upon the previous literature upon the influence of reward upon visual behaviour, one can ask whether eye movement can be considered a type of nichepicking within the visual scene, and whether this is a type of emotional regulation. In the next section, we explore literature that has investigated this angle, and seek to explore whether individuals subconsciously avoid looking at non-rewarding stimuli, whilst actively attending to stimuli they find rewarding.

### 1.1.5 Emotion, Preference and Eye Movements

Eye-tracking studies of clinical cohorts provide evidence for the outcomes that may occur when this (i.e. visual nichepicking) potential form of emotional regulation goes awry. A recent meta-analytic review by Armstrong and Olatunji (2012) synthesised 33 experiments that investigated eye movement indicators for depression and anxiety. They found individuals with depression spend more time fixating upon dysphoric stimuli than controls, and conversely demonstrate reduced orientation towards, and dwell time upon, positive stimuli. This 'reversing' of attentional bias demonstrates that the results are not attributable to a global reduction in attending to stimuli, but a specific bias towards identifying and focusing upon items with a negative emotional salience (Armstrong & Olatunji, 2012). This supports the proposed framework by suggesting an interplay between defective emotional regulation and the allocation of visual attention (e.g., orientating and maintenance of gaze) toward affective stimuli. Although, as eye movement behaviour is heavily influenced by task and reward, it should be noted that this may alternatively be described by a motivation to seek unfavourable information (Cisler & Olatunji, 2012; Giesler, Josephs, & Swann, 1996).

An investigation into the relationship between gaze and subjective interpretations of facial attractiveness found that gaze plays an active role in preference formation (Shimojo, Simion, Shimojo, & Scheier, 2003). Shimojo et al. (2003) describe a preference-specific mechanism, where the mere exposure effect (Harmon-Jones & Allen, 2001) and the general tendency for neurotypical individuals to look at stimuli that they prefer (Glaholt & Reingold, 2009; Schotter et al., 2010) form a positive feedback loop that enhances the cortical representation of that region; aiding its selection for a response before the decision is consciously made (Simion & Shimojo, 2006). This is often referred to as 'gaze bias', and provides an example of eye movements indicating personal preference. This becomes of direct interest when one views the task of visual exploration within the context of a reinforcement learning paradigm - our third and final theme of interest (Hayhoe, Mennie, Gorgos, Semrau, & Sullivan, 2004).

In reinforcement learning, we consider the task of exploring the environment and making actions that maximise the cumulative reward of these actions (Lecun, Bengio, & Hinton, 2015). We have two main factors to consider: whether to exploit existing knowledge of an action ('exploitation') or to explore the environment in search of more rewarding behaviours ('exploration'). As human behaviour takes place in a constantly changing environment, direct analogies between reinforcement learning and the task of whether to continue fixating upon a particular item ('exploit') or move to sample a different portion of the visual scene ('explore') can be made. In a similar way to how dopaminergic activity represents a bidirectional prediction error signal in the substantia nigra (Chang et al., 2015), recent evidence suggests that the Locus Coeruleus (LC) norepinephrine (NE) system plays a direct role in supporting the expression of exploration or exploitation strategies (Aston-Jones & Cohen, 2005). The LC-NE system tends towards two types of activity: phasic and tonic. Aston-Jones, Rajkowski, and Cohen (1999) recorded neuronal activity within the LC whilst primates conducted a visual search task, where they were rewarded with juice for correct responses. They found that the responses of LC neurons were selective, with strong phasic responses corresponding to presentations of target stimuli, whilst distractor stimuli elicited much weaker or no response (e.g. tonic firing). Furthermore, these neuronal responses preceded the behavioural responses by  200 milliseconds (Aston-Jones et al., 1999). From this research, phasic activity is proposed to emerge when task-congruent outcomes are observed (the expressed behaviour has 'high utility'), and promote continued engagement with this behaviour ('exploitation'). Tonic activity is associated with disengagement from the current task and engagement with exploration behaviour. Due to the LC-NE having widespread afferent connections to motor and decision-making regions of the brain (Aston-Jones et al., 1999), this provides a distributed influence that acts as a temporal attention filter to enhance the expression of task-relevant behaviours when reward is high, and promote exploration otherwise (Aston-Jones & Cohen, 2005).

This final mechanism allows visual behaviour to be dominated by the expression of task-relevant 'just in time' eye movements, provided that engaging with the task is rewarding to the individual, and to be more exploratory otherwise (Aston-Jones & Cohen, 2005). When combined with insights from the gaze cascade literature of Shimojo et al. (2003) and the nichepicking literature of Campos et al. (2004), this suggests that we may conceptualise task-free visual behaviour in particular as being guided (in part, as visual attention is still influenced by low-level features; Itti & Koch, 2001) by past rewarding experiences which encapsulate the individual's personal preferences. However, I note that the magnitude of this association may rely upon an appropriate visual environment; in the same way that an individual can not attend a theme park (e.g., display nichepicking behaviour) if it is shut, the opportunity for an individual to display visual nichepicking may also be dependant upon the opportunities provided by the visual scene. As such, to strongly elicit differences in visual nichepicking that may be informative of personal attributes, it may be essential to present a stimulus within which the individual can select and pigeonhole their visual attention in such a way that will indicate their preferences.

Overall, from the point of view of predicting a broad range of personal attributes from visual behaviour, the literature outlined within this segment suggests that a range of attributes (e.g. individual variance reflecting the individual's personal preferences and learnt associations) are most likely to be reflected within visual behaviour during the exploration of a visual stimulus that provides plenty of opportunities for the individual to express visual nichepicking. Additionally, engaging in a rewarding task may reduce the amount of individual variance expressed that is due to the individual's preferences, due to salience being heavily dominated by task relevance in this situation (Aston-Jones & Cohen, 2005). However, I note that this does not imply that a rewarding task will lead to individuals exhibiting identical behaviour. Individual differences in visual behaviour may still occur, but are likely to reflect prior experience or expertise (e.g., as exhibited within expert batsmen; Land & McLeod, 2000) rather than alternative psychological factors which are the focus of this thesis.

### 1.1.6 Eye Movements and Attention: Summary

In this section, I have explained how eye movement events such as fixations and saccades provide a documentation of the individual's approach to exploring and exploiting information within the visual scene. A wide range of literature has been presented to highlight that where we look is influenced by saliency, and that saliency is, in part, dictated by factors that are not physical properties of the scene. This includes the task being performed and, most importantly to this thesis, the individual's learnt associations with objects (or regions/contexts) within the visual scene. The converging theme is that eye movements are influenced by the individual's personal attributes (e.g. what they perceive to be rewarding), and I have suggested that employing a free viewing design within a stimulus that offers a wide variety of social and affective content may provide an ideal platform for the expression of visual behaviour that reflects personal attributes. As such, the next section explores which personal attributes have been found to be reflected in eye movements, and the current state of the art in predicting personal attributes from visual behaviour.

**Eye movements and Personal Attributes**

In the previous sections, I have highlighted the importance of task and stimulus properties in eliciting eye movements that reflect the individual's associations with objects or regions within the scene. However, there is literature that suggests that even visual stimuli that lack emotional or social cues may still reflect aspects of an individual's personality. In a recent investigation into individual differences in eye movements, Bargary et al. (2017) collected data from 1058 participants whilst they completed four visual tasks. Two of the tasks were designed to elicit saccades and fixations (a pro and anti-saccade task), whilst the remaining tasks sought to elicit smooth pursuit behaviour (tracking a slowly moving dot that either maintained a set trajectory or moved randomly upon the screen). None of these tasks involved stimuli with social or affective properties. The authors also collected a range of questionnaire based outcomes, including basic demographic information (e.g., sex, age), a short measure of personality (Donnellan, Oswald, Baird, & Lucas, 2006) under the Big Five personality construct (Costa, 1996), and an autism quotient scale (Baron-Cohen, Wheelwright, Skinner, Martin, & Clubley, 2001). From this, the authors identified significant sex differences in how the participants tracked a moving target with their eyes and the speed of initiating a saccade towards a target (for an introduction to pro-saccadic tasks, see: Munoz & Everling, 2004). Namely, Bargary et al. (2017) found that males were significantly better at accurately tracking a smoothly moving object (females were more likely to make a saccade that took them further from the target than the previously fixated location) and were significantly faster at initiating saccades in response to an onset signal. Furthermore, after controlling for sex differences, there was a significant (but weak: Spearman's $p = .15$) association between the personality trait of Extroversion, which is related to how social and outgoing an individual is, and worse performance upon the smooth pursuit task. No significant association was found between an individual's eye movements and autism quotient score. This suggests that oculomotor behaviour may be informative of sex and specific aspects of the individual's personality (i.e. Extroversion) within tasks that involve smooth pursuit behaviour even when the stimulus is lacking in social and affective properties. This is surprising as trait Extroversion is conceptually linked to the individual's affinity for social situations, and thus should not influence an individual's oculomotor behaviour when social cues are absent in the stimulus. It is not well known why individuals produce distinctive eye movements in this case, but it has been speculated that cognitive differences in decision making regions may may feed back to lower levels of visual processing to contribute to these differences (Bargary et al., 2017). As such, this literature suggests that we may conceptualise eye movements as being deeply ingrained and entwined with the individual's psyche, to the point that their individual preferences for sociability are subconsciously (i.e. automatically) expressed even when exploring visual scenes that contain no trait-congruent stimuli. However, I note that the carefully controlled experimental settings required to elicit these differences in visual behaviour, combined with the weak nature of the associations, may limit the potential for practical application within ecological settings. Furthermore, there is no clear rationale why some personal attributes (e.g., gender and Extroversion) are expressed in the visual behaviour captured by Bargary et al. (2017) but not others. This is a gap in the literature which requires future research to be fully understood.

To provide a brief overview of the big five theory of personality, it has been proposed that there are five main aspects of an individual's personality, and that these aspects influence how the individual thinks and behaves across a broad range of situations (Costa, 1996; McCrae & Costa, 2004). These five personality traits are Openness (with high scorers being more open to new experiences), Conscientiousness (high scorers being more conscientious), Extroversion (high scorers being more sociable), Agreeableness, and Neuroticism (high scores being more anxious and neurotic). An individual's score upon these five traits (e.g. being low or high) is relatively stable across time (McCrae & Costa, 2004), and it has been demonstrated that knowledge of an individual's big five personality traits can be useful for targeting them with persuasive material (this topic is covered in depth in later sections, but see: Matz, Kosinski, Nave, & Stillwell, 2017; Matz & Netzer, 2017, for further details).

The research of Rauthmann, Seubert, Sachse, and Furtner (2012) also indicates that visual stimuli that lack emotional or social cues may still reflect aspects of an individual's personality. The authors showed two abstract animations (simulated cellular structures), one characterised by fast jerky movements (coloured red), and another with slow smooth movements (coloured blue) to a cohort of 242 undergraduate participants. Using a hierarchical linear modelling technique, they investigated the association between the big five personality traits and three oculomotor metrics (mean fixation duration, total dwell time, and number of fixations). The authors also investigated the Behavioral Inhibition System, and Behavioral Approach/Activation System (Carver & White, 1994) which are linked to the big five traits of Neuroticism and Extroversion respectively (Keiser & Ross, 2011). Overall, the big five personality traits were able to significantly improve the ability to predict the three oculomotor metrics compared to knowing only which stimulus was being presented. This demonstrates that there is a significant association between an individual's personality profile and their oculomotor behaviour. However, the results were weak, with the Big Five personality scores accounting for at most 4% of variance in the oculomotor metric ($R^2 \leq .04$), and Openness was the only individual Big Five trait to be significantly associated with any oculomotor metric (associated with longer mean fixation duration and dwell times; Rauthmann et al., 2012).

The two studies presented above are both well powered, featuring large sample sizes (1058 and 242 participants respectively; Bargary et al., 2017; Rauthmann et al., 2012). Both illustrate that weak linear associations between eye movements and the individual's personal attributes (e.g., personality and sex) emerge within experimental conditions. However, as demonstrated by Harrison, Binetti, Coutrot, Johnston, and Mareschal (2018), this association does not occur in all stimuli and/or visual tasks. The authors tracked the eye movements of 77 participants whilst they watched video clips featuring actors looking directly at the camera (e.g. making direct eye content with the viewer). The direct gaze duration varied per video clip, and for each video clip the participant was asked to indicate whether the direct gaze of the actor felt 'too short' or 'too long'. From this data, the authors were able to calculate each individual's preferred gaze duration and characterise the participant's visual behaviour via total fixation duration upon the actors eye region, and the number of blinks. After collecting the participant's big five personality trait information via the NEO-Five-Factor Inventory (NEO-FFI) 60 item questionnaire (McCrae & Costa, 2004), the authors found no significant association between scores upon any of the personality traits and the participant's subjective preference for gaze duration (as assessed via the rating task) or measures of visual behaviour (fixation duration upon the actors eye region/number of blinks). This was contrary to expectations, as given the social nature of the stimuli, the authors expected personality traits such as Extroversion (linked to sociability) and Openness (to experience) to be reflected within these visual metrics (Harrison et al., 2018).

Given that the above literature employs tightly controlled experimental paradigms, a key question that remains is how much the individual may disclose via their eye movements when viewing a naturalistic stimulus which contains a wide array of content and social contexts. For example, an extrovert may find a social scene more interesting (i.e. rewarding/motivating) than an introvert, leading to the extrovert 'exploiting' this content for longer than the introvert. As such, measuring eye movements in such an environment may increase the signal to noise ratio (as suggested in the previous section), leading to the ability to better predict personal attributes (e.g., the big five personality traits) from visual behaviour. However, it appears that not all eye movements exhibit linear associations with personal attributes, as demonstrated by Harrison et al. (2018). A direction for my thesis is to identify the extent to which personal attributes are decodable from everyday visual behaviour (using a wide range of description metrics), such as whilst browsing social media. This section illustrated that weak linear associations may occur between visual behaviour and personal attributes such as age and personality traits. In the next section, I move on to introduce the current state of the art in employing non-linear techniques to predict personal attributes from visual behaviour, and explain how this is relevant within the context of everyday behaviour.

## 1.2 The potential for classifying personal traits and attributes from eye movements

Eye-based Human Computer Interaction (eb-HCI) has arrived, with an ever expanding range of devices from manufacturers that allow customers to interact with their laptop or desktop computer using oculomotor movements (Chen & Hennessey, 2018). Such features are now natively supported by major operating systems such as Windows 10, which allows users to move the cursor, type and improve text-to-speech functions using eye movements (Coldewey, 2020). This hints towards the potential for widespread adoption as the costs of hardware and barriers to entry continue to reduce. However, we note that such designs are in their infancy, and currently significant barriers exist to incorporating eb-HCI into our everyday computer interactions. Gaze based interactions are less precise than traditional keyboard and mouse interfaces, which leads to applications often requiring re-designing to be gaze-interaction friendly (e.g. by enlargement of buttons and control surfaces). For example, Chen and Hennessey (2018) found that completing a simple email composition task using a keyboard and mouse took participants on average 33.73 seconds with a zero error rate, whilst participants conducting the same task whilst engaging with an gaze input interface took on average 422.6 seconds with eight errors. Optimising this interface to accommodate the reduced precision of eb-HCI improved the average time to completion to 232 seconds with three errors. As such, a key barrier to the widespread uptake of eb-HCI (beyond the additional costs to the consumer) is the requirement that application developers redesign user interfaces to accommodate for the reduced precision provided by this interaction method.

Previous sections of this review have demonstrated that weak linear associations exist between visual behaviour and latent user features such as personality or age. To recap, eye movements are considered a correlate of attention that indirectly reflects the user's preferences, associations, and biases (Itti & Koch, 2001; Niu et al., 2012; Shimojo et al., 2003). Perhaps more importantly, eye movements give insight into how an individual reaches a behavioural outcome (Hayhoe et al., 2004; Hayhoe & Ballard, 2005). This unique insight into the human mind offers an unprecedented granularity of information in comparison to observations of digital outcomes (e.g. mouse clicks and keyboard input) alone. For instance, digital footprints may comprise behavioural outcomes such as clicking 'like' on a particular selection of Facebook pages. In contrast, eye movement data will provide an account of exactly what on the page was attended to, and thus likely consciously perceived, in the lead up to each of these behavioural outcomes. With commercially available (e.g. marketed to non-research audiences) eye tracking devices, such as the Tobii Eye Tracker Five (Gaming, 2020) that is capable of recording at 133hz, data can now be collected that reflects cognitive processes at a rate of 133 data points a second. Across many users, this very quickly enters the domain of big data, with a substantial increase in the velocity and volume of data available compared to the digital footprints previously discussed. This raises several questions regarding what may be possible with this additional insight into a user's decision process; could the new type of targeted advertising wait for your eye patterns to indicate a particular mood before showing you their product? Should this raise ethical concerns when that product is a political ideology?

Whilst speculation raises pertinent questions about the future of privacy law, it is baseless without empirical evidence. As such, this next segment seeks to identify which attributes can currently be predicted from eye movement data when applying more advanced (often non-linear) methods. A powerful method of capturing complex patterns of associations is to apply supervised machine learning, where a set of features (e.g. eye movement descriptions) with corresponding outcomes (e.g., labelled 'male' or 'female') are used to train an algorithm how to predict the outcome given the features. The benefit of supervised machine learning paradigms is that many algorithms exist that are able to learn complicated non-linear interactions between the input and output, and the performance of the model is tested upon examples that the algorithm has never seen before. In the context of this thesis, this provides a direct index of what attributes can be predicted from an individual, given a set of descriptions about their eye movement behaviour. For the interested reader, a full discussion of the benefits and drawbacks of applying a machine learning approach to prediction over traditional statistical techniques in psychology is given in the methods section.

Tseng et al. (2013) demonstrated that the eye movements of participants passively watching a stream of video snippets for 15 minutes provides sufficient information for a machine learning algorithm (support-vector machine using recursive feature elimination) to discriminate between clinical cohorts, and between clinical cohorts and controls. They recorded the eye movements of individuals with Parkinson's disease (PD), attention deficit hyperactivity disorder (ADHD), and fetal alcohol spectrum disorders (FASD). Each represents a disorder of the underlying mechanisms that drive oculomotor behaviour. In PD patients, disruption within the basal ganglia leads to over inhibition of voluntary movement; including oculomotor control. ADHD is characterised by a short attention span, which is again reflected in the allocation of visual attention and thus oculomotor behaviour. In contrast, FASD has an underlying cause (maternal consumption of alcohol) that leads to widespread cortical changes (e.g., reduced brain volume; Riley & McGee, 2005). This often leads to deficits in attentional allocation, which in oculomotor terms represents a similar but mechanistically different disorder to ADHD (Tseng et al., 2013).

Using the oculomotor data gained from recording participants viewing a 15 minute stream of video snippets (each snippet lasting 2-4 seconds), the authors derived a total of 224 descriptions of eye movements (henceforth referred to as features; Tseng et al., 2013). These features were split into three distinct groups; 48 features describing the oculomotor events themselves ('oculomotor'), 160 saliency based metrics ('saliency') that describe actual gaze allocation compared to predictions from a biologically inspired model of scene saliency, and 16 group-based features ('group') that reflected deviations in participants' gaze allocation compared to a normative group of young adult controls. A separate model was built for each classification task. In discriminating between PD patients (N=14) and age matched controls (N=24), comparing models built separately upon each feature group, 'oculomotor' features provided the best overall model with 86.4% accuracy (chance: 63.2%). Using a label permutation test, it was found that eye movement features describing the distribution and parameters of oculomotor events, such as the peak velocity, duration, and amplitude of saccades, significantly improved the model's performance compared to chance (Tseng et al., 2013). The model built using the 'group' features performed second best, with 74.6% accuracy (significantly above chance), and finally the model derived from 'saliency' features was not significantly better than chance. The model built using all features performed slightly better than the oculomotor feature only model, with 89.6% accuracy. This suggests that features describing the oculomotor events themselves, rather than comparisons to statistical models or other groups, hold the most relevant information toward the goal of discriminating between PD and age matched controls. However, including additional non-oculomotor features does provide a modest 3% gain in overall accuracy. Using this technique, the authors were also able to predict significantly better than chance whether an individual belongs to the ADHD (N=21) or control group (N=18), where the discrepancy between predicted saliency and participant gaze behaviour formed the most important feature group (Tseng et al., 2013), and between individuals with a FASD (N=13) and controls (N=18), with a mixture of salience and group features being most important

(Tseng et al., 2013).

This work suggests that a range of different neurological conditions are able to be predicted through collecting the eye movements of an individual passively viewing video streams. These neurological conditions have a direct impact upon how the individual expresses visual behaviour and are a form of sensitive personal information. As such, the literature is relevant to the aim of this thesis, which seeks to understand whether eye movements may lead to the individual unknowingly disclosing personal information. However, this thesis is focused upon the use of eye movements whilst participants browse social media to predict non-clinical personal information. Of note is that the input used was specifically designed to highlight bottom-up visual mechanisms, as would not be the case when viewing social media. As such, two main questions remain unanswered: whether predictions of clinical psychological attributes can be made when the stimuli vary across participants, and if predictions about personality traits rather than clinical conditions can be made from oculomotor features. Both of these questions represent a more difficult (or subtle) problem than addressed in previous literature. For instance, having the stimuli vary across participants means that differences between participants in eye movements may be due to external factors such as low-level scene saliency rather than individual differences, which may obscure eye movement patterns that are linked to the outcome of interest (e.g., as the signal to noise ratio decreases). Furthermore, clinically diagnosed neurological conditions by definition reflect an underlying psychobiological dysfunction that results in social deviance or conflicts with society (Psychiatric Association, 2013). In contrast, sub-clinical personality attributes do not have to meet this criteria, and therefore are likely to have less impact upon the individual's everyday behaviour. Because of this, it may be that personality attributes are more subtly expressed within eye movements than clinical disorders - making the prediction of such traits a more difficult task.

Regarding whether private (i.e. not explicitly disclosed) traits are predictable from the eye movements of individuals interacting with naturalistic stimuli that vary across participants (e.g. not purposefully made to elicit attribute-congruent visual behaviour), a useful example comes from Rello and Ballesteros (2015), who compiled (and thus re-purposed) 1,135 eye tracking recordings of individuals with and without dyslexia reading various pieces of text. In pursuit of a dyslexia pre-screening tool, this was used to build a binary classifier which achieved 80% accuracy in distinguishing between the two groups. This demonstrates that when the task is uniform across participants, data can be aggregated and successfully used to derive useful predictive models for an implicit, if still clinical, personal attributes (even when the stimulus varies across participants). Together with data from Tseng et al. (2013), this suggests a range of clinical conditions can be predicted from oculomotor behaviour with relatively high accuracy, and that models can be derived through eye movements in response to naturalistic stimuli.

To build upon the previous points, the work of Hoppe, Loetscher, Morey, and Bulling (2018) illustrates that even when the stimulus varies relatively dramatically across participants (i.e., eye movements outdoors within diverse naturalistic visual scenes), oculomotor behaviour can be used to form predictive models of non-clinical personal attributes. Participants (N=42) were instructed to walk around for 10 minutes and visit the local shop whilst having their eye movements tracked at 60hz using a head-mounted eye tracker. Upon returning, participants completed a questionnaire assessing the big five personality traits (60-item NEO-FFI; Costa, 1996). Once the data was collected, participants were evenly allocated into three groups representing low, medium, and high scorers for that trait. These three labels form the classification outcomes within a supervised learning setting. The authors derived 207 features from the eye tracking data. They took three key approaches to creating features from visual behaviour. One approach was to divide the visual scene into an eight by eight grid and calculate oculomotor metrics (e.g. total fixation duration) in correspondence to the grid. As this is within the participant's frame of reference, this provides an index of where the participant directed their visual attention from their own point of view (e.g. did they often look up and to the right?). The authors also characterised the statistical features of oculomotor events, such as the average fixation duration or saccade amplitude. Finally, the authors created "*n*-gram" features by recording the frequency of occurrence of patterns of fixations and saccades. To elaborate, the authors defined each saccade or fixation as being either large or small, and then calculated the frequency with which *n* events occurred in sequence (e.g., a 3-gram might be: small fixation, small saccade, large fixation). For each personality trait, a random forest classifier was built using all features, with each feature being standardised such that the mean was zero and the standard deviation was one. A sliding window approach was used to ensure gaze data was independent from viewing duration, and preserve individual differences in viewing behaviours (Bulling, Ward, Gellersen, & Tröster, 2011). Several different lengths of sliding window were computed, and cross-validation (for details on how cross validation works, see methodology chapter section 2.4.3) was used to find the optimal window length for each trait. Hoppe et al. (2018) found that, for the labels of neuroticism (40.3%), extroversion (48.6%), agreeableness (45.9%) and conscientiousness (43.1%) the classifier performed significantly above chance. The performance of these models is modest compared to the accuracy of clinical classifications presented above (e.g., see: Rello & Ballesteros, 2015; Tseng et al., 2013), but this is to be expected considering the more subtle pattern of attentional allocation. This provides evidence that personality can be predicted from oculomotor data, and highlights that modest gains can be established from eye movement data obtained using low-resolution recording equipment.

The research of Berkovsky et al. (2019) builds upon the work of Hoppe et al. (2018) by demonstrating that describing visual behaviour in response to a series of images chosen for their affective properties can lead to strong improvements in classifier accuracy for non-clinical personal attributes. Berkovsky et al. (2019) selected 50 images from the international affective picture system (Mikels et al., 2005) by considering two features: the valence (positive/negative) and arousal state (low/high) of the image. For each of the four resulting combinations (i.e., high / low arousal and positive / negative valence), they selected ten images. The authors also selected ten images that were neutral upon both attributes, leading to a total of 50 images. Each set of ten images is referred to as a block, as they were presented together (e.g., all high valence, positive images were shown in sequence together). The participant's eye position whilst viewing the images was recorded using a head-mounted eye tracker, recording at 60hz. Participants also completed a range of personality measures via questionnaire using the HEXACO personality framework (an alternative to the big five theory; see: Lee & Ashton, 2004), which formed the outcome variables to be predicted. As with previous literature, the authors framed the problem as a supervised three category classification paradigm, with instances allocated equally into the low, medium, and high categories. For each of the five blocks, the authors created ten variables (50 features in total measured) that describe the participants' visual behaviours, including: fixations (average number and duration); saccades (average number, velocity, and amplitude, and peak amplitude); pupil size (horizontal and vertical diameter); number of blinks; saccade-fixation ratio. As such, the above features describe content-based physiological behaviour in response to a block of images with similar valence and arousal ratings. After conducting feature selection using correlation-based methods, the personality traits of openness (85%), conscientiousness (80.95%), extroversion (80.95%), agreeableness (90.48%) and resiliency (80.95%, equivalent to neuroticism under the big five theory; Lee & Ashton, 2004) were predicted with high accuracy. Such strong results are perhaps surprising given that past literature has generally struggled to achieve results that are substantially above chance (e.g., as in Hoppe et al., 2018), and warrant further inspection. An initial reason why these results should be interpreted with caution is because they are based upon a very small cohort of only 21 participants, which provides fewer than seven observations for the algorithm to learn from per category (i.e., low, medium, high). This is too few to evaluate whether the findings will generalise well to the general population (this relates to the curse of dimensionality; see, Alpaydin, 2014). Furthermore, a critical review of the eye movement features used reveals that many may be redundant or inaccurate. For example, two measures treated as independent variables are pupil width and height respectively; these two measures are without doubt redundant, as the pupil contracts and dilates in a circular manner (e.g., in accordance to a circular sphincter muscle). Perhaps most concerning is the potential for artefactual differences between the two measures to occur (Brisson et al., 2013), with systemic measurement errors being well documented within eye tracking systems with twice the sampling rate of that used within Berkovsky et al. (2019). This is as using camera-based pupil size estimation leads to distortion of the pupil as the participant moves their head and eyes in relation to the camera (Brisson et al., 2013). Furthermore, the saccadic peak amplitude (the fastest speed reached during a saccade) is particularly sensitive to the recording method (Mack, Belfanti,

& Schwarz, 2017), linearly related to the saccadic amplitude, and only reliable when the sampling rate is above 200hz (Schmitt, Muser, Lanz, Walz, & Schwarz, 2007). As such, this is likely to produce variance in at least three of the ten metrics, which is attributable to measurement error rather than participant behaviour. Within such a small sample, it is possible that the pattern recognition algorithms found spurious associations between the input (e.g. oculomotor metrics) and outcome variables that may not generalise to other eye tracking systems or cohorts. However, this research does provide preliminary evidence that a wide range of non-clinical attributes may be decodable from visual behaviour using pattern recognition techniques.

Overall, a wide range of research has illustrated that both clinical and non-clinical attributes can be predicted from an individual's visual behaviour when more powerful pattern recognition techniques (such as supervised machine learning) are applied. This has been found within both experimental settings (Berkovsky et al., 2019; Tseng et al., 2013) and everyday visual stimuli (Hoppe et al., 2018; Rello & Ballesteros, 2015). However, with the growing presence and ubiquity of eye based human computer interaction interfaces, there is still a clear gap in the literature regarding whether personal attributes can be predicted whilst browsing common online websites which contain a broad range of social and affective contexts. In the next section, I explore which types of personal attribute may be displayed within online settings that meet the above criteria.

## 1.3 Predicting Personal Attributes from Online Behaviour

To recap, the dominant model of personality in psychology is the "big five" model, which rates individuals along five scales: Openness, Agreeableness, Conscientiousness, Extroversion, and Neuroticism ('OCEAN': Egan, Deary, & Austin, 2000). This has traditionally been assessed via questionnaires, a methodological approach that has been extensively investigated and found to return good inter-sample reliability and consistency (Sartori, 2010). Additionally, it has recently been found that estimates of an individual's personality score can be established from their online activity (Eftekhar, Fullwood, & Morris, 2014; Gosling, Augustine, Vazire, Holtzman, & Gaddis, 2011; Kosinski, Stillwell, & Graepel, 2013). In a meta-analysis, Azucar, Marengo, and Settanni (2018) identified 774 studies which attempted to predict personality from digital footprints, of which 14 investigated independent data sets and provided adequate methodological details, and thus were included in the meta-analysis. Results found the accuracy of prediction (evaluated upon instances not used to train the machine learning algorithm) was consistent across all five personality traits, with a meta-correlation ranging between 0.29 and 0.4; these correlations generally increased in strength when multiple records of digital footprints were included in the overall model (Azucar et al., 2018). Examples of investigated digital footprints are the type of language used in text-based submissions to social media sites or patterns of interaction with content, such as 'liking' a Facebook post. The results of Azucar et al. (2018) suggests that different records of digital footprints either add value by representing non-overlapping novel information about an individual's personality or (as suggested by the authors) by representing interactions between the different types of digital footprints. This demonstrates that if a user provides their behavioural data (e.g. digital footprints) to a social media provider, a range of predictive methodologies can be used to implicitly assume attributes about the user (which the user may not have realised when they gave consent for their behaviour to be tracked).

An example of predicting personality from digital footprints is the finding that an individual's pattern of 'likes' upon the social media platform Facebook has been found to accurately predict a range of sensitive personal attributes (Kosinski et al., 2013). This includes their alignment with political ideologies (Republican/Democrat = 85% accuracy), gender (93% accuracy), and sexuality (Homo/Heterosexual = 88% accuracy). The above binary classification models were created by identifying 55,814 possible different items that a user could 'like', and then reducing this to 100 features using singular value decomposition (this helped avoid overfitting, where the model performs well upon the training data but generalises poorly to new unseen instances; see the methods section for details). This demonstrates that sensitive information can be inferred from online behaviour without an individual's explicit consent (e.g., Facebook users did not explicitly state that they consented for their pattern of 'liking' to be used in this way) to a high degree of accuracy.

A wide variety of literature focusing upon the prediction of user personality from social media site information has emerged over the past decade, and today the ability to predict personal attributes from digital records now outperforms human intuition. Youyou, Kosinski, and Stillwell (2015) found that, compared to human predictions of an individual's personality (based upon the participant's friend rating them using the ten-item IPIP scale of Donnellan et al., 2006), predictions based upon Facebook 'likes' displayed higher accuracy and external validity when predicting lifestyle outcomes such as substance use and physical health.

Furthermore, researchers are able to reach very wide audiences with online studies. For example, Ortigosa, Carro, and Quiroga (2014) illustrates how researchers may create Facebook applications that collect personal data within a viral marketing campaign to reach large cohorts. Using such an approach, Ortigosa et al. (2014) reached a Spanish cohort of 65,000 users in four weeks, and used this data to build a machine learning model predicting the users' personalities using the alternative five personality model (a less common personality construct similar to the big five; see, Aluja et al., 2006). Data was collected regarding the participants' friends, frequency of posts, and activity upon their page. The authors created independent models for each trait, with labels distributed evenly across three bins (low, medium, high). Resultant models predicted each of the personality traits above chance, with around 70% accuracy (Ortigosa et al., 2014). This illustrates that Facebook 'likes' are not uniquely useful, and that other social information obtained from a user's social media profile may be useful for predicting personality. However, I note that a lack of detail regarding how the model's performance was evaluated and likely sampling bias from the data collection method ('viral' distribution from the authors' Facebook account) requires that results from Ortigosa et al. (2014) be interpreted with caution.

An insight into the potential for this type of research to be used in commercial settings comes from a recent paper investigating the impact of persuasive media when it was tailored to an individual's personality (Matz et al., 2017). This technique of displaying advertisements based upon an individual's (inferred) psychological profile is described as 'psychological persuasion' by Matz et al. (2017) and refers to the principle that, based upon a psychological trait, a persuasive stimulus may be effective for one individual but ineffective for another. In this study, the participant's (N= 3.5 million) Openness and Extroversion category was inferred from records of their digital behaviour (utilising the participant's 'likes' upon the Facebook platform, imitating the approach previously demonstrated by Kosinski et al., 2013), after which the authors presented the participant with an advertisement (a beauty product or mobile game application) that was marketed in a way as to be either congruent or incongruent with their personality profile. The authors found that matching the persuasive material to the individual's psychological profile led to significantly increased click-through and conversion rates. The effect was most pronounced for low-openness advertisements aimed at congruent Facebook users, with users within this group being 1.79 times as likely to install the promoted application. As such, from the established associations between 'liking' behaviour on Facebook and personality trait scores, Matz et al. (2017) were able to use Facebook's advertising framework to identify individuals who were likely to score high (or low) upon Openness and Extroversion and successfully target them with tailored content.

The above literature illustrates the commercial value of digital records, as it seems that psychological persuasion is an effective method of influencing an individual's behaviour, with the caveat of being demonstrated only for consumer behaviour concerning arguably trivial items (beauty products and mobile game applications). However, one must keep in mind the scale of social media sites; Facebook has 1.4 billion daily active users on average (Kosinski, Matz, Gosling, Popov, & Stillwell, 2015), and the study by Matz et al. (2017) reached 3.5 million social media accounts. This provides a mechanism by which interested parties may exert influence (via targeted advertisements) upon a widespread audience, the potential implications of which are discussed in the next section.

### 1.3.1 The potential risks to the individual and society that may emerge from implicit disclosure of personal information via eye tracking technology

The potential for mass persuasion has not been ignored by political parties, with reports that the leave campaign for the 'brexit' referendum spent £3.5 million on targeted social media adverts (The Electoral Commission, 2018); this is, perhaps, overshadowed by the improper acquisition and use of Facebook data to psychologically persuade voters (Committee, 2019), which the above literature would suggest has empirical merit as an effective method of mass persuasion (see: Matz et al., 2017). Regardless of the validity of such allegations, it is clear that psychological persuasion techniques have the potential to influence human behaviour, have caught the eye of interested parties, and can be applied across a large enough scale to concern the reader about societal impact. Furthermore, the way in which personal attributes can be predicted through digital footprints is growing. For example, Kosinski and Wang (2017) employed deep neural nets to extract features from a large database of facial images and then applied logistic regression upon these features for classification of sexual orientation. The final model reached 81% accuracy in distinguishing between gay and heterosexual males, compared to the 61% accuracy of human judges. This suggests that even publishing a photograph of your face online is enough for implicit assumptions to be made about your preferences (Kosinski & Wang, 2017).

Whilst the potential for unethical conduct (e.g., the undemocratic manipulation of voting behaviour) that accompanies predicting private attributes from online behaviour (including potentially from eye movements) raises concern, there is equally scope for positive impact. Aside from clear applications in public health campaigns for aiding the persuasion of individuals to take positive actions, there is also research into using social media information to inform interventions for depression and mental illness (Roberts, Kuncel, Shiner, Caspi, & Goldberg, 2007). Additionally, it is possible for users to exert a measure of control over what inferences can be derived about them. Chen, Fraiberger, Moakler, and Provost (2016) demonstrate that withholding just a few online activities is enough to drastically reduce what can be reliably inferred, and propose a 'cloaking device' that can help users do so. However, it is noted that such countermeasures are not infallible and can be circumvented by retraining the predictive model (Chen et al., 2016).

## 1.4 Conclusion

To conclude this introduction, I have reviewed a wide range of factors that influence how individuals distribute their visual attention across a scene. This includes the concept of visual saliency (Hayhoe & Ballard, 2005; Itti & Koch, 2001), the importance of task (Tatler et al., 2010), and how reward teaches unique patterns of visual behaviour (Anderson, 2013; Chelazzi et al., 2013; Hayhoe, 2000; Kowler, 2011). I have used these concepts to highlight that the visual salience of a given stimulus is a combination of its physical properties, task-relevance, and the affective associations held by the individual to the stimulus. This is because unique aspects of what an individual finds rewarding or adverse (unfavourable) in the visual environment shapes how the individual learns to deploy visual routines in relation to that task, where 'reward' is the achievement of task-congruent actions (Anderson, 2013). Together, this provides a framework that explains how the allocation of visual attention can be biased by non-physical properties (e.g., the affective associations the individual holds towards items within the scene; Anderson, 2013) and highlights that we can capture these biases via eye tracking techniques. This concept is supported by a broad range of literature illustrating that individual variance in visual behaviour (as reflected in the spatio-temporal pattern of oculomotor events such as fixations and saccades) across the visual scene is informative of a broad range of personal attributes (e.g, Bargary et al., 2017; Rauthmann et al., 2012). I then expanded upon this literature to provide an overview of the state of the art in predicting personal attributes from visual behaviour to illustrate that the production of oculomotor behaviour may reflect both clinical attributes (Rello & Ballesteros, 2015; Tseng et al., 2013) and sub-clinical traits (Berkovsky et al., 2019; Hoppe et al., 2018). Key to this literature is that much stronger performance can be obtained when applying machine learning techniques (e.g., Azucar et al., 2018), compared to results from traditional linear models (e.g., Rauthmann et al., 2012). Finally, I have outlined the potential risks to the individual's privacy (Chen et al., 2016; Matz et al., 2017) and society (Parliament Commons Select Committee, 2018) that may emerge from the implicit disclosure of personal information via eye tracking technology. In particular, I highlight that the volume and temporal granularity of oculomotor data is much greater in comparison to the types of digital footprints commonly used for psychological profiling (e.g., clicking 'like' or interacting with content; Settanni, Azucar, & Marengo, 2018), which may provide the opportunity for greater invasions of privacy than has previously been possible.

As such, having demonstrated the plausibility of predicting personal attributes from visual behaviour, and the potential risk to privacy and society engendered by such an approach, the goal of this thesis is to explore which private information, if any, can be reliably implied from participant viewing behavior whilst browsing upon a social media website. In this way, I assess the potential risk to a user's privacy that may stem from incorporating eye based interaction into everyday computer activity. Given the above literature and objectives, a major part of this thesis will be the collection of eye movement data and the application of machine learning techniques to this data. Both topics are addressed in detail in the following methodology chapter.

## References

Alexander, G. E., DeLong, M. R., & Strick, P. L. (1986). Parallel Organization of Functionally Segregated Circuits Linking Basal Ganglia and Cortex. *Annu. Rev. Neurosci. 9*(1), 357–381. doi:10.1146/annurev.ne.09.030186.002041

Alpaydin, E. (2014). *Introduction to machine learning*. MIT press.

Aluja, A., Rossier, J., García, L. F., Angleitner, A., Kuhlman, M., & Zuckerman, M. (2006). A cross-cultural shortened form of the ZKPQ (ZKPQ-50-cc) adapted to English, French, German, and Spanish languages. *Pers. Individ. Dif. 41*(4), 619–628. doi:10.1016/j.paid.2006.03.001

Anderson, B. (2013). A value-driven mechanism of attentional selection. *J. Vis. 13*(3), 7–7. doi:10.1167/13.3.7

Armstrong, T., & Olatunji, B. O. (2012). Eye tracking of attention in the affective disorders: A meta-analytic review and synthesis. *Clin. Psychol. Rev. 32*(8), 704–723. doi:10.1016/j.cpr.2012.09.004

Aston-Jones, G., & Cohen, J. D. (2005). An integrative theory of locus coeruleus-norepinephrine function: Adaptive Gain and Optimal Performance. *Annu. Rev. Neurosci. 28*(1), 403–450. doi:10.1146/annurev.neuro.28.061604.135709

Aston-Jones, G., Rajkowski, J., & Cohen, J. (1999). Role of locus coeruleus in attention and behavioral flexibility. *Biol. Psychiatry*, *46*(9), 1309–1320. doi:10.1016/S0006-3223(99)00140-7

Azucar, D., Marengo, D., & Settanni, M. (2018). Predicting the Big 5 personality traits from digital footprints on social media: A meta-analysis. *Pers. Individ. Dif. 124*, 150–159. doi:10.1016/j.paid.2017.12.018

Bargary, G., Bosten, J. M., Goodbourn, P. T., Lawrance-Owen, A. J., Hogg, R. E., & Mollon, J. (2017). Individual differences in human eye movements: An oculomotor signature? *Vision Res. 141*, 157–169. doi:10.1016/j.visres.2017.03.001

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The Autism-Spectrum Quotient (AQ): Evidence from Asperger Syndrome/High-Functioning Autism, Malesand Females, Scientists and Mathematicians. *Journal of Autism and Developmental Disorders*, *31*(1), 5–17. doi:10.1023/A:1005653411471

Battista, J., Kalloniatis, M., & Metha, A. (2005). Visual function: The problem with eccentricity. *Clin. Exp. Optom. 88*(5), 313–321. doi:10.1111/j.1444-0938.2005.tb06715.x

Berkovsky, S., Taib, R., Koprinska, I., Wang, E., Zeng, Y., Li, J., & Kleitman, S. (2019). Detecting Personality Traits Using Eye-Tracking Data. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (pp. 1–12). CHI '19. doi:10.1145/3290605.3300451

Bhorkar, G. (2017). A Survey of Augmented Reality Navigation. *Presence Teleoperators Virtual Environ. 6*(4), 355–385. doi:10.1561/1100000049

Brisson, J., Mainville, M., Mailloux, D., Beaulieu, C., Serres, J., & Sirois, S. (2013). Pupil diameter measurement errors as a function of gaze direction in corneal reflection eyetrackers. *Behavior Research Methods*, *45*(4), 1322–1331. doi:10.3758/s13428-013-0327-0

Bulling, A., Ward, J. A., Gellersen, H., & Tröster, G. (2011). Eye movement analysis for activity recognition using electrooculography. *IEEE Trans. Pattern Anal. Mach. Intell. 33*(4), 741–753. doi:10.1109/TPAMI.2010.86

Campos, J. J., Frankel, C. B., & Camras, L. (2004). On the nature of emotion regulation. *Child Dev. 75*(2), 377–394. doi:10.1111/j.1467-8624.2004.00681.x

Carrasco, M. (2011). Visual attention: The past 25 years. *Vision Research*, *51*(13), 1484–1525. doi:10.1016/j.visres.2011.04.012

Carver, S, & White, L. (1994). Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment : The BIS/BAS scales. *Journal of Personality and Social Psychology*, *67*(2), 319–333.

Chang, C. Y., Esber, G. R., Marrero-Garcia, Y., Yau, H. J., Bonci, A., & Schoenbaum, G. (2015). Brief optogenetic inhibition of dopamine neurons mimics endogenous negative reward prediction errors. *Nat. Neurosci. 19*(1), 111–116. doi:10.1038/nn.4191

Chelazzi, L., Perlato, A., Santandrea, E., & Della Libera, C. (2013). Rewards teach visual selective attention. *Vision Res. 85*, 58–62. doi:10.1016/j.visres.2012.12.005

Chen, C. W., & Hennessey, C. (2018). Online Eye-gaze Usability Evaluation of Gmail: Are mobile interfaces easier to use with eye-trackers? *C. Proc. 33*(1), 2–4.

Chen, D., Fraiberger, S. P., Moakler, R., & Provost, F. (2016). Enhancing Transparency and Control when Drawing Data-Driven Inferences about Individuals. *Big data*, *5*(3), 197–212. doi:10.1089/big.2017.0074

Chen, N., Bi, T., Zhou, T., Li, S., Liu, Z., & Fang, F. (2015). Sharpened cortical tuning and enhanced cortico-cortical communication contribute to the long-term neural mechanisms of visual motion perceptual learning. *NeuroImage*, *115*, 17–29. doi:10.1016/j.neuroimage.2015.04.041

Cisler, J. M., & Olatunji, B. O. (2012). *Emotion regulation and anxiety disorders*. doi:10.1007/s11920-012-0262-2

Cohen, M. R., & Newsome, W. T. (2009). Estimates of the Contribution of Single Neurons to Perception Depend on Timescale and Noise Correlation. *J. Neurosci. 29*(20), 6635–6648. doi:10.1523/JNEUROSCI.5179-08.2009

Coldewey, D. (2020). Facebook, YouTube, Netflix and more get eye-tracking apps from Tobii.

Committee, C. S. (2019). *Disinformation and 'fake news': Final Report*. The Digital, Culture, Media and Sport Committee.

Costa, P. T. (1996). Work and personality: Use of the NEO-PI-R in industrial/organisational psychology. *Appl. Psychol. 45*(3), 225–241. doi:10 . 1111 / j . 1464 - 0597 . 1996 . tb00766.x

De Valois, R, & De Valois, K. (1980). Spatial vision. *Annu. Rev. Psychol. 31*(1), 309–341. doi:10.1146/annurev.ps.31.020180.001521

Donnellan, M. B., Oswald, F. L., Baird, B. M., & Lucas, R. E. (2006). The mini-IPIP scales: Tiny-yet-effective measures of the Big Five factors of personality. *Psychological Assessment*, *18*(2), 192–203. doi:10.1037/1040-3590.18.2.192

Eftekhar, A., Fullwood, C., & Morris, N. (2014). Capturing personality from Facebook photos and photo-related activities: How much exposure do you need? *Comput. Human Behav. 37*, 162–170. doi:10.1016/j.chb.2014.04.048

Egan, V., Deary, I., & Austin, E. (2000). The NEO-FFI: Emerging British norms and an item-level analysis suggest N, A and C are more reliable than O and E. *Pers. Individ. Dif. 29*(5), 907–920. doi:10.1016/S0191-8869(99)00242-1

Emery, N. J. (2000). The eyes have it: The neuroethology, function and evolution of social gaze. *Neurosci. Biobehav. Rev. 24*(6), 581–604. doi:10.1016/S0149-7634(00)00025-7

Funahashi, S. (2006). Prefrontal cortex and working memory processes. *Neuroscience*, *139*(1), 251–261. doi:10.1016/j.neuroscience.2005.07.003

Gaming, T. (2020). Tobii Eye Tracker 5 - The Next Generation of Head and Eye Tracking. https://gaming.tobii.com/product/eye-tracker-5/.

Giesler, R. B., Josephs, R. A., & Swann, W. B. (1996). Self-verification in clinical depression: The desire for negative evaluation. *J. Abnorm. Psychol. 105*(3), 358–368. doi:10.1037//0021-843X.105.3.358

Glaholt, M. G., & Reingold, E. M. (2009). Stimulus exposure and gaze bias: A further test of the gaze cascade model. *Attention, Perception, Psychophys. 71*(3), 445–450. doi:10.3758/APP.71.3.445

Gollisch, T., & Meister, M. (2010). Eye Smarter than Scientists Believed: Neural Computations in Circuits of the Retina. *Neuron*, *65*(2), 150–164. doi:10.1016/j.neuron.2009.12.009

Gosling, S. D., Augustine, A. A., Vazire, S., Holtzman, N., & Gaddis, S. (2011). Manifestations of Personality in Online Social Networks: Self-Reported Facebook-Related Behaviors and Observable Profile Information. *Cyberpsychology, Behav. Soc. Netw. 14*(9), 483–488. doi:10.1089/cyber.2010.0087

Hamamé, C. M., Cosmelli, D., Henriquez, R., & Aboitiz, F. (2011). Neural Mechanisms of Human Perceptual Learning: Electrophysiological Evidence for a Two-Stage Process. *PLOS ONE*, *6*(4), e19221. doi:10.1371/journal.pone.0019221

Harmon-Jones, E., & Allen, J. J. B. (2001). The Role of Affect in the Mere Exposure Effect: Evidence from Psychophysiological and Individual Differences Approaches. *Personal. Soc. Psychol. Bull. 27*(7), 889–898. doi:10.1177/0146167201277011

Harrison, C., Binetti, N., Coutrot, A., Johnston, A., & Mareschal, I. (2018). Personality Traits Do Not Predict How We Look at Faces: *Perception*. doi:10.1177/0301006618788754

Hayhoe, M, Mennie, N, Gorgos, K, Semrau, J, & Sullivan, B. (2004). The Role of Internal Models and Prediction in Catching Balls. *J. Vis. 4*(8), 156–156. doi:10.1167/4.8.156

Hayhoe, M. (2000). Vision using routines: A functional account of vision. *Vis. cogn. 7*(1-3), 43–64. doi:10.1080/135062800394676

Hayhoe, M., & Ballard, D. (2005). Eye movements in natural behavior. *Trends Cogn. Sci. 9*(4), 188–194. doi:10.1016/j.tics.2005.02.009

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & van de Weijer, J. (2011). *Eye Tracking: A Comprehensive Guide To Methods And Measures*.

Hoppe, S., Loetscher, T., Morey, S. A., & Bulling, A. (2018). Eye Movements During Everyday Behavior Predict Personality Traits. *Front. Hum. Neurosci. 12*, 105. doi:10.3389/FNHUM.2018.00105

Ikemoto, S., Yang, C., & Tan, A. (2015). Basal ganglia circuit loops, dopamine and motivation: A review and enquiry. *Behav. Brain Res. 290*, 17–31. doi:10.1016/j.bbr.2015.04.018

Ilango, A., Kesner, A. J., Keller, K. L., Stuber, G. D., Bonci, A., & Ikemoto, S. (2014). Similar roles of substantia nigra and ventral tegmental dopamine neurons in reward and aversion. *J. Neurosci. 34*(3), 817–22. doi:10.1523/JNEUROSCI.1703-13.2014

Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nat. Rev. Neurosci. 2*(3), 194–203. doi:10.1038/35058500

Joshua, M., Adler, A., Mitelman, R., Vaadia, E., & Bergman, H. (2008). Midbrain Dopaminergic Neurons and Striatal Cholinergic Interneurons Encode the Difference between Reward and Aversive Events at Different Epochs of Probabilistic Classical Conditioning Trials. *J. Neurosci. 28*(45), 11673–11684. doi:10.1523/JNEUROSCI.3839-08.2008

Keiser, H. N., & Ross, S. R. (2011). Carver and whites' bis/fffs/bas scales and domains and facets of the five factor model of personality. *Personality and Individual Differences*, *51*(1), 39–44. doi:10.1016/j.paid.2011.03.007

Kelley, T. A., & Yantis, S. (2010). Neural Correlates of Learning to Attend. *Front. Hum. Neurosci. 4*. doi:10.3389/fnhum.2010.00216

Kosinski, M., & Wang, W. (2017). Deep Neural Networks Are More Accurate Than Humans at Detecting Sexual Orientation From Facial Images | Stanford Graduate School of Business. *J. Personal. Soc. Psychol. (in Press. 114*(2), 246–257.

Kosinski, M., Stillwell, D., & Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. *Proc. Natl. Acad. Sci. 110*(15), 5802–5805. doi:10.1073/pnas.1218772110

Kosinski, M., Matz, S. C., Gosling, S. D., Popov, V., & Stillwell, D. (2015). Facebook as a research tool for the social sciences: Opportunities, challenges, ethical considerations, and practical guidelines. *Am. Psychol. 70*(6), 543–556. doi:10.1037/a0039210

Kowler, E. (2011). Eye movements: The past 25years. *Vision Res. 51*(13), 1457–1483. doi:10.1016/j.visres.2010.12.014

Kredel, R., Vater, C., Klostermann, A., & Hossner, E.-J. (2017). Eye-Tracking Technology and the Dynamics of Natural Gaze Behavior in Sports: A Systematic Review of 40 Years of Research. *Front. Psychol. 8*, 1845. doi:10.3389/fpsyg.2017.01845

Land, M., & McLeod, P. (2000). From eye movements to actions: How Batsmen hit the ball. *Nat. Neurosci. 3*(12), 1340–1345. doi:10.1038/81887

Lecun, Y., Bengio, Y., & Hinton, G. (2015). *Deep learning*. doi:10.1038/nature14539

Lee, K., & Ashton, M. C. (2004). Psychometric Properties of the HEXACO Personality Inventory. *Multivariate Behavioral Research*, *39*(2), 329–358. doi:10.1207/s15327906mbr3902_8

Liu, P., & Basso, M. A. (2008). Substantia Nigra Stimulation Influences Monkey Superior Colliculus Neuronal Activity Bilaterally. *J. Neurophysiol. 100*(2), 1098–1112. doi:10.1152/jn.01043.2007

Mack, D. J., Belfanti, S., & Schwarz, U. (2017). The effect of sampling rate and low-pass filters on saccades – A modeling approach. *Behavior Research Methods*, *49*(6), 2146–2162. doi:10.3758/s13428-016-0848-4

Mai, J. K., Majtanik, M., & Paxinos, G. (2015). *Atlas of the human brain*. Academic Press Inc.

Matz, S. C., Kosinski, M., Nave, G., & Stillwell, D. J. (2017). Psychological targeting as an effective approach to digital mass persuasion. *Proceedings of the National Academy of Sciences*, *114*(48), 12714–12719. doi:10.1073/pnas.1710966114. eprint: https://www.pnas.org/content/114/48/12714.full.pdf

Matz, S. C., & Netzer, O. (2017). Using Big Data as a window into consumers' psychology. *Current Opinion in Behavioral Sciences*, *18*, 7–12. doi:10.1016/j.cobeha.2017.05.009

Maunsell, J. H. R., & Treue, S. (2006). Feature-based attention in visual cortex. *Trends in Neurosciences*. Neural Substrates of Cognition, *29*(6), 317–322. doi:10.1016/j.tins.2006.04.001

McCrae, R., & Costa, P. (2004). A contemplated revision of the NEO Five-Factor Inventory. *Pers. Individ. Dif. 36*(3), 587–596. doi:10.1016/S0191-8869(03)00118-1

McGinty, V. B., Rangel, A., & Newsome, W. T. (2016). Orbitofrontal Cortex Value Signals Depend on Fixation Location during Free Viewing. *Neuron*, *90*(6), 1299–1311. doi:10.1016/j.neuron.2016.04.045

Mikels, J. A., Fredrickson, B. L., Larkin, G. R., Lindberg, C. M., Maglio, S. J., & Reuter-Lorenz, P. A. (2005). Emotional category data on images from the international affective picture system. *Behavior Research Methods*, *37*(4), 626–630. doi:10.3758/BF03192732

Milgram, P., & Drascic, D. (2001). Perceptual Effects In Aligning Virtual And Real Objects In Augmented Reality Displays.

Montague, P. R., Hyman, S. E., & Cohen, J. D. (2004). Computational roles for dopamine in behavioural control. *Nature*, *431*(7010), 760–767. doi:10.1038/nature03015

Munoz, D. P., & Everling, S. (2004). *Look away: The anti-saccade task and the voluntary control of eye movement*. doi:10.1038/nrn1345

Niu, Y., Todd, R. M., & Anderson, A. K. (2012). Affective salience can reverse the effects of stimulus-driven salience on eye movements in complex scenes. *Front. Psychol. 3*(SEP). doi:10.3389/fpsyg.2012.00336

Ortigosa, A., Carro, R. M., & Quiroga, J. I. (2014). Predicting user personality by mining social interactions in Facebook. In *J. Comput. Syst. Sci.* (Vol. 80, pp. 57–71). doi:10.1016/j.jcss.2013.03.008

Parliament Commons Select Committee. (2018). *Evidence from Christopher Wylie, Cambridge Analytica whistle-blower, published - News from Parliament - UK Parliament*.

Pirenne, M. H. (1962). Visual Acuity. In H. Davson (Ed.), *The Visual Process* (pp. 175–195). doi:10.1016/B978-1-4832-3089-4.50018-2

Psychiatric Association, A. (2013). *The Diagnostic and Statistical Manual of Mental Disorders* (5th ed.). Washington, DC.

Ramnani, N., & Owen, A. M. (2004). *Anterior prefrontal cortex: Insights into function from anatomy and neuroimaging*. doi:10.1038/nrn1343

Rauthmann, J. F., Seubert, C. T., Sachse, P., & Furtner, M. R. (2012). Eyes as windows to the soul: Gazing behavior is related to personality. *J. Res. Pers. 46*(2), 147–156. doi:10.1016/j.jrp.2011.12.010

Rello, L., & Ballesteros, M. (2015). Detecting readers with dyslexia using machine learning with eye tracking measures. In *Proc. 12th Web All Conf. - W4A '15* (pp. 1–8). doi:10.1145/2745555.2746644

Riley, E., & McGee, C. (2005). Fetal Alcohol Spectrum Disorders: An Overview with Emphasis on Changes in Brain and Behavior. *Exp. Biol. Med. 230*(6), 357–365. doi:10.1177/15353702-0323006-03

Roberts, B. W., Kuncel, N. R., Shiner, R., Caspi, A., & Goldberg, L. R. (2007). The Power of Personality: The Comparative Validity of Personality Traits, Socioeconomic Status, and Cognitive Ability for Predicting Important Life Outcomes. *Perspect. Psychol. Sci. 2*(4), 313–345. doi:10.1111/j.1745-6916.2007.00047.x

Robinson, D. A. (1972). Eye movements evoked by collicular stimulation in the alert monkey. *Vision Res. 12*(11), 1795–1808. doi:10.1016/0042-6989(72)90070-3

Robles, E., Laurell, E., & Baier, H. (2014). The retinal projectome reveals brain-area-specific visual representations generated by ganglion cell diversity. *Curr. Biol. 24*(18), 2085–2096. doi:10.1016/j.cub.2014.07.080

Saito, T., Nouchi, R., Kinjo, H., & Kawashima, R. (2017). Gaze bias in preference judgments by younger and older adults. *Front. Aging Neurosci. 9*(AUG), 285. doi:10.3389/fnagi.2017.00285

Sartori, R. (2010). Face validity in personality tests: Psychometric instruments and projective techniques in comparison. *Qual. Quant. 44*(4), 749–759. doi:10.1007/s11135-009-9224-0

Schmitt, K.-U., Muser, M. H., Lanz, C., Walz, F., & Schwarz, U. (2007). Comparing eye movements recorded by search coil and infrared eye tracking. *Journal of Clinical Monitoring and Computing*, *21*(1), 49–53. doi:10.1007/s10877-006-9057-5

Schotter, E. R., Berry, R. W., McKenzie, C. R. M., & Rayner, K. (2010). Gaze bias: Selective encoding and liking effects. *Vis. cogn. 18*(8), 1113–1132. doi:10.1080/13506281003668900

Serences, J. T., & Yantis, S. (2006). *Selective visual attention and perceptual coherence*. doi:10.1016/j.tics.2005.11.008

Settanni, M., Azucar, D., & Marengo, D. (2018). Predicting Individual Characteristics from Digital Traces on Social Media: A Meta-Analysis. *Cyberpsychology, Behav. Soc. Netw. 21*(4), 217–228. doi:10.1089/cyber.2017.0384

Shimojo, S., Simion, C., Shimojo, E., & Scheier, C. (2003). Gaze bias both reflects and influences preference. *Nat. Neurosci. 6*(12), 1317–1322. doi:10.1038/nn1150

Simion, C., & Shimojo, S. (2006). Early interactions between orienting, visual sampling and decision making in facial preference. *Vision Res. 46*(20), 3331–3335. doi:10.1016/j.visres.2006.04.019

Simons, D. J., & Chabris, C. F. (1999). Gorillas in our midst: Sustained inattentional blindness for dynamic events. *Perception*, *28*(9), 1059–1074. doi:10.1068/p281059

Summerfield, C., & Egner, T. (2009). Expectation (and attention) in visual cognition. *Trends Cogn. Sci. 13*(9), 403–409. doi:10.1016/j.tics.2009.06.003

Tatler, B., Baddeley, R., & Gilchrist, I. (2005). Visual correlates of fixation selection: Effects of scale and time. *Vision Res. 45*(5), 643–659. doi:10.1016/j.visres.2004.09.017

Tatler, B., Hayhoe, M., Land, M., & Ballard, D. (2011). Eye guidance in natural vision: Reinterpreting salience. *J. Vis. 11*(5), 5–5. doi:10.1167/11.5.5

Tatler, B. W., Wade, N. J., Kwan, H., Findlay, J. M., & Velichkovsky, B. M. (2010). Yarbus, eye movements, and vision. *Iperception. 1*(1), 7–27. doi:10.1068/i0382

The Electoral Commission. (2018). *Report on the 23 June 2016 referendum on the UK's membership of the European Union*.

Treue, S. (2003). Visual attention: The where, what, how and why of saliency. *Curr. Opin. Neurobiol. 13*(4), 428–432. doi:10.1016/S0959-4388(03)00105-3

Tseng, P.-H., Cameron, I. G. M., Pari, G., Reynolds, J. N., Munoz, D. P., & Itti, L. (2013). High-throughput classification of clinical populations from natural viewing eye movements. *J. Neurol. 260*(1), 275–284. doi:10.1007/s00415-012-6631-2

Volkmann, F. C., Schick, A. M., & Riggs, L. A. (1968). Time course of visual inhibition during voluntary saccades. *J. Opt. Soc. Am. 58*(4), 562–569. doi:10.1364/JOSA.58.000562

Vuilleumier, P, Henson, R. N., Driver, J, Dolan, R. J., & Orban, G. A. (2002). Multiple levels of visual object constancy revealed by event-related fMRI of repetition priming. *Nat. Neurosci. 5*(5), 491–9. doi:10.1038/nn839

Youyou, W., Kosinski, M., & Stillwell, D. (2015). Computer-based personality judgments are more accurate than those made by humans. *Proc. Natl. Acad. Sci. 112*(4), 1036–1040. doi:10.1073/pnas.1418680112

# Chapter 2

# Methodological Approaches: Eye tracking and Machine Learning

In this chapter, I summarise the key methods used to address the main research goal of the thesis, which is to explore whether personal attributes (e.g., personality trait information) may be predicted from the eye movements of users whilst they browse online social networking site (SNS) content. As discussed in the introduction, I address this question because eye movements are informative of a wide range of clinical (Rello & Ballesteros, 2015; Tseng et al., 2013) and non-clinical personal attributes (Rauthmann, Seubert, Sachse, & Furtner, 2012), as they reflect the way attention can be driven by our learned associations with objects within the scene (Anderson, 2013).

In my first study, I used a linear regression analysis to evaluate whether three visual metrics (total fixation duration, number of fixations, and time to first fixation) that have been found to correlate with preference evaluation in static image-based paradigms (i.e., as found in: Holmes & Zanker, 2012; Maughan, Gutnikov, & Stevens, 2007; Nummenmaa, Hyönä, & Calvo, 2006) also reflect preference decisions whilst participants view a mock SNS. This first study provides insight into whether the additional features present in SNS environments (e.g., scrolling to occluded items) influences the association between visual behaviour and the participant's subjective evaluation of content within the visual scene. The scrolling design required careful stimulus design and counterbalancing to avoid order effects, which is described in this chapter.

As reviewed within the introduction, machine learning techniques are particularly capable of producing strong predictive models from eye movement data and can outperform traditional linear techniques. As such, using the eye movement data obtained from the first study, in my second study I employ a machine learning approach whereby a broad array of eye movement descriptions are used to see if it is possible to predict a new outcome that was collected but not analysed previously: the participant's personality trait category (low or high) under the big five theory of personality (Costa, 1996). This study further utilises the counterbalanced design employed in study one to investigate whether descriptions of visual behaviour in response to either the content (i.e., a given SNS post) or location (i.e., the top most SNS post) leads to the best predictive performance when used to train a machine learning algorithm to predict the individual's personality trait category.

As the counterbalanced design used in the first two studies does not accurately reflect how participants may browse SNS content in an everyday, natural setting, in the third study I recorded participants browsing their own Facebook news feed content. This leads to the content of the page varying across participants (as it is personalised to each user), and I describe the steps taken to allow for a meaningful comparison of visual behaviour across participants in this chapter. As such, the third study explores whether personality traits can be predicted under these more naturalistic conditions. In my fourth and final study, I explore the utility of using segmentation techniques to transform personality scores into categorical outcomes, as this is an essential step for any researcher wishing to apply the classification algorithms used within this thesis. In particular, I evaluated whether applying more advanced segmentation strategies influenced classifier performance. I then used the segmentation methodology developed in the first half of study four to evaluate if new outcomes (such as the participant's age and self-esteem) could be predicted from their eye movements. This fourth study seeks to broaden the range of personal attributes known to be predictable from eye movement behaviour upon SNS content, whilst providing a rigorous framework for future work wishing to transform continuous questionnaire values into discrete categories.

This concludes the main research questions addressed in this thesis. The next section provides an overview of the methodology employed to answer these questions, after which I deliver deeper into explaining how the methodology has been applied in each of my studies.

## 2.1 Methodology Employed

The research questions outlined above frequently required that I measure and create descriptions of eye movement behaviour relative to content or spatial location on the screen. In the first two studies this involved the creation and presentation of naturalistic scrolling webpages that emulate SNS content. The third study required the hand-coding of social media content into categories, which are used again to describe eye movement behaviour in the fourth and final study.

Alongside measuring and describing the eye movement of participants, I collected a range of measures (outcome variables) regarding the participants' personal attributes. In study one, this included preference ratings for different examples of content, whilst in studies two and three I measured the participants personality trait scores under the big five theory (Costa, 1996). Demographic variables (e.g., age, sex) were collected and reported in all studies and investigated directly in study four alongside psychological traits such as narcissism and self-esteem. Finally, in study four, I employed unsupervised learning techniques upon the questionnaire results (personality, self-esteem, and narcissism) to segment participants into categorical groups for each psychological trait. Having armed the reader with a broad overview of the methodological demands accompanying the studies reported within this thesis, I will now move on to provide more details.

## 2.2 Eye tracking methodology

In this section, I describe how eye movement data was collected and processed in order to describe how participants explored the visual scene. As discussed within the introduction, the expression of eye movements is influenced by situational factors, including the demands of the current task (Tatler, Wade, Kwan, Findlay, & Velichkovsky, 2010) and the surrounding visual environment (Itti & Koch, 2001). As a key research outcome for this thesis is to contribute an insight into the risk posed by allowing eye tracking devices to be enabled in everyday situations, it is important that the participant feels comfortable enough to display naturalistic behaviour (i.e., equivalent to that which they would display at home whilst browsing online content). This is to ensure that our results have ecological validity and are applicable to the main research question.

To allow participants to engage with browsing tasks in a naturalistic manner and support the ecological validity of our tasks, all recordings took place within an ambiently lit room using an infrared eye tracker (Tobii TX300). This device allows for some free movement of the head (as opposed to requiring a chin rest), and records samples at 300hz. Importantly, the infrared light does not distort the visual scene and is reflected by both the pupil and cornea. The two reflections can be tracked through computer vision algorithms, and by calculating the varying position of the pupil relative to the more stable corneal reflection as the participant views known locations, the system can interpolate to provide a model of the relationship between these two reflections at a set distance; this is termed calibration. I used the in-built Tobii five-point calibration sequence in all studies. This calibration is valid for the situation in which it was formed, and deviations in viewing distance, lighting, and the subject's position from this state may render subsequent recordings unreliable (Holmqvist et al., 2011). Additionally, it is common for recording quality to 'drift' over time, making it advisable for experiments of longer duration to re-calibrate as necessary (Holmqvist, Nyström, & Mulvey, 2012). As repeated recalibration of the eye tracker is likely to break participant immersion within the task, within each study, I keep the eye tracking task short enough to not require re-calibration, and I visually inspect the recordings to guard against meaningful drift in recording quality. For example, in the first data set collected, the eye tracking task is three minutes long, and in the second only the first twenty seconds of the recording is used. Participants were also seated in a sturdy desk chair without castors, which limited movement to prevent large disturbances to the viewing distance.

Having established a model of the individual participant's eye, the eye tracker can then report the sampled horizontal and vertical positions of the pupil over time relative to the plane of the computer screen. A filter algorithm can then be applied to the data to detect fixations and saccades (and other eye movement types may be classified if required; Andersson, Larsson, Holmqvist, Stridh, & Nyström, 2017). This results in the acquisition of a sequence of oculomotor events, accompanied by their attributes. Examples include the duration of fixations, along with the amplitude and duration of saccades. In the studies within this thesis, I utilise an established velocity threshold based fixation filter (I-VT filter) with a window length of 20ms and a velocity threshold of 30°/s to segment the stream of eye coordinates into the events of fixations and saccades. This choice of filter was selected as velocity based filters with these parameters have been found to perform well in high-resolution eye tracking settings and are considered the standard in the field (Holmqvist et al., 2011; Stuart et al., 2019). To confirm that realistic oculomotor behaviour had been captured, I inspected visualised movies of saccade/fixation patterns within the Tobii Studio software.

This choice of fixation filter has implications when dealing with smooth pursuit behaviour. This is because during smooth pursuit, the velocity threshold of the filter is unlikely to be exceeded, which leads to the smooth pursuit being counted as a long fixation located at the point of origin. In this thesis, the visual stimuli are restricted to Facebook style social media webpages. In this context, the only likely causes of smooth pursuit will be either from the individual tracking the same location as they scroll (in which case, the representation as a long fixation is still valid - they are attending to that location), or within video-based stimuli embedded within a post. In the latter case, the video will be contained in a single region of interest, which ensures that the characterisation of the visual behaviour is still correct. As such, on balance, this drawback is not likely to be of practical concern within our paradigm.

Finally, to compare differences in spatially specific oculomotor behaviour across participants, in each study, I label the scene with areas of interest (AOI), with each label being represented as a range of spatial coordinates within the visual scene (Hessels, Kemner, van den Boomen, & Hooge, 2016). This can be viewed as a 'binning' approach to evaluating eye behaviour, as the detected oculomotor events can be classed as either within ('hit') or outside ('miss') the spatial coordinates of interest. Furthermore, statistics, such as how long it takes before a fixation is made within a given region (time to first fixation), can be calculated using this methodology.

### 2.2.1 Feature Engineering: Describing eye movements

Having described how I tracked the eye movements of participants, I now move on to explain how I calculated metrics which represent the spatial and temporal distribution of eye movements across the given stimulus. It is important to note that this section is designed to provide an overview; not all of these metrics are used in each study, and the subset of visual metrics used for each study is detailed in later sections. As discussed earlier, AOIs provide information about the spatial dispersion of eye movement behaviour by allowing us to ask 'when and how much did this region attract attention'. Note that this also provides temporal information, allowing us to ask 'how long did it take until the participant fixated here' (i.e., time to first fixation). As such, it is possible to calculate the time to first fixation, total fixation duration, number of fixations, and the number of visits (how many times the participant fixated inside the AOI, fixated outside, then returned to the AOI) in relation to each defined AOI (Duchowski, 2017).

**Metrics calculated by AOI**

- Time to First Fixation
- Total Fixation Duration
- Number of Fixations
- Number of Visits

Statistical descriptions can be calculated to characterise (i.e., describe the distribution of) the properties of the eye movement events (i.e., fixations and saccades) over a given time. In this thesis, I consider fixation events to have two properties and saccades to have three.

**Fixation Properties**

- Duration - 'How long did they fixate each time'
- Frequency - 'How many fixations did they make'

**Saccade Properties**

- Duration - 'How long did each saccade take'
- Amplitude - 'What was the angular distance covered each time'
- Frequency - 'How many saccades did they make'

I calculated statistical descriptions, such as the mean, standard deviation, and range, for the duration and amplitude properties. This allowed the creation of a broad set of metrics by which to describe the participants oculomotor behaviour, which were not linked to any AOI. It is important to note that the above statistics (i.e., mean, standard deviation and range) are examples, and that the set of statistical functions applied varied by study; details are provided in the study-specific sections below.

## 2.3 Outcome Variables: Collecting personal information

Assessing the ability of classifiers to predict aspects of the individual's private attributes required that I first collect a range of personal attributes from the participant. This begins with basic demographic information, such as the participant's age, sex, and race. I also collected information about the participant's voting behaviour ('did you vote in the last election') and political preference via an analogue slider scale from 'extreme left' (-1) to 'extreme right' (+1). This information was collected from all participants. In specific studies, additional information was collected via the questionnaires and materials described below.

### 2.3.1 Subjective Stimulus Ratings

To collect the subjective rating of each social media post in the mock SNS webpage (study one), each participant responded to the question 'How did viewing this image make you feel?'. The image (social media post content) in question was then presented with a 100 point analogue slider centered on zero, spanning -1 ('negative') and +1 ('positive'), directly below. All five social media posts were rated.

### 2.3.2 Personality

Following previous literature (e.g., Hoppe, Loetscher, Morey, & Bulling, 2018; Rauthmann et al., 2012), I evaluated personality under the big five theory (Costa, 1996) using the NEO-FFI 60 item scale (McCrae & Costa, 2004). This scale was chosen as it provides a relatively quick measure of the big five personality traits, whilst being more precise and having greater test-retest reliability (Cronbach's Alpha .88; McCrae & Costa, 2004) than other shorter options (e.g., the ten item personality scale of Gosling, Rentfrow, & Swann, 2003, which has a Cronbach's Alpha of .68).

There are five personality traits under the big five theory (Costa, 1996), and the NEO-FFI 60 item scale provides a score between zero (lowest) and 48 (highest) for each trait. To allow for the application of more powerful pattern recognition techniques (i.e., in a classification paradigm; Alpaydin, 2014), I created categorical outcomes for each of these continuous traits. I checked that the categorical outcomes represent significantly different groups with independent t-tests. In the second study, I applied a median split technique to equally divide the participant's into low and high categories for each personality trait. In the third study I employ a quantile-based strategy to equally divide participants into low, medium, and high categories for each personality trait; this was chosen to explore whether the paradigm could provide a greater granularity of insight into the personality of the participant. As before, I check for significant differences between groups. The median and quantile split techniques were chosen as they have the advantage of preventing a large class imbalance, which ensures there are enough examples of each category for the classifier to learn from. These techniques have also been used in previous literature (e.g., Berkovsky et al., 2019; Hoppe et al., 2018).

### 2.3.3 Self-Esteem

I collected an index of the participant's self-esteem (study four) via the ten item Rosenberg Self-Esteem Scale (Robins, Hendin, & Trzesniewski, 2001). The Rosenberg Self-Esteem Scale was selected as it exhibits high internal consistency (Cronbach's Alpha .88 to .90; Robins et al., 2001) and it is well established within the literature. The scale returns a maximum score of 40 (high self esteem) and a minimum score of 10 (low self esteem).

### 2.3.4 Narcissism

The participant's narcissism score was evaluated using the Narcissistic Personality Inventory (NPI) 16 item scale (Ames, Rose, & Anderson, 2006). This short measure provides similar internal, discriminant and predictive validity to the longer 40 item NPI scale (Ames et al., 2006), and was selected to provide a brief account of the individual's narcissism score. The scale returns a maximum score of 16 (high narcissism), and a minimum of zero (low narcissism).

## 2.4 Machine Learning Methods

Here, it is worth reflecting upon the benefits and drawbacks of employing a machine learning approach in comparison to methods traditionally employed within psychological research. Most machine learning algorithms are robust outside of parametric assumptions; the key requirement is that the samples are independent and identically distributed (Alpaydin, 2014). The machine learning paradigm is best employed when the researcher is focused upon the accurate prediction of outcomes, is not being constrained by biological plausibility, and comes at the expense of interpretability (i.e., the resultant model may not be interpretable by the human mind; Yarkoni & Westfall, 2017). Accordingly, machine learning methods can provide better predictive performance, but require that the researcher shifts their focus from identifying casual mechanisms to understanding whether prediction is possible at all (and to what extent).

Machine learning models are evaluated upon observations that they have never encountered before ('out of sample'), thus the performance reported reflects the model's ability to generalise outside of the data used to train it. In the classification paradigm considered within this thesis, machine learning algorithms are concerned with learning a function that maps a set of features ($X$) to an outcome ($y$) within the feature space defined by $X$. As this thesis is focused upon how well a researcher could predict the private attributes of participants (e.g., the outcome usually reflects a questionnaire answer or score $y$) from their visual behaviour (i.e., a set of features consisting of the eye tracking metrics $X$), adopting a machine learning based methodological approach is an ideal fit; by doing so, I am not constraining the analysis to models that are readily interpreted in terms of causal mechanisms, and instead concern myself primarily with out-of-sample performance, relating to the question of how accurately personal information eye movement data can be predicted. As such, I employ a machine learning approach in the following studies as it provides an excellent way to identify statistical patterns in the data that are useful for predicting the given outcome. When considering which oculomotor metrics to provide the model (i.e. when forming the set of features $X$), I had to balance a trade-off between two competing objectives. The first objective is to provide enough information for the machine learning algorithm to learn from, so that it is able to distinguish between classes. The second objective is to avoid including too many variables, from which spurious correlations may be identified that lead to the classifier performing poorly upon observations outside of the sample it was trained upon. The solution I implemented for this was to break down a large corpus of features inspired by previous literature (covered in detail in chapter-specific sections) into different sets, and evaluate one set at a time. This limits the number of variables considered at once and reduces the number of observations required to populate the feature group $X$ (For the interested reader, see 'the curse of dimensionality' as described in Alpaydin, 2014), whilst providing an additional granularity of insight into which sets of features are most informative of a given outcome.

In machine learning paradigms, the out-of-sample performance of the model upon a given evaluation metric (e.g., accuracy) is evaluated through a process called cross-validation, where the data is split into training and testing sets. The two cross-validation techniques applied within this thesis are Leave-One-Out (LOO) and k-fold. Both are common within the machine learning literature, with LOO being generally a safe choice that provides an optimal approximation of the prediction error, with the smallest variance for a fixed model (Zhang & Yang, 2015), and k-fold providing a robust and computationally efficient manner of evaluating prediction error within large data sets (Alpaydin, 2014; Raeder, Hoens, & Chawla, 2010). Each will be described in the following sections to highlight the insight that each approach affords the researcher. After this, I introduce the concept of nested cross-validation as a way of reducing the amount of data required to train and evaluate the model. As a note on terminology, in this section I use 'algorithm' to refer to the mechanism used to learn from the data, whilst a 'model' is the resulting function that maps the input (i.e., features) to the output. A classifier is a specific type of model that outputs categorical predictions.

### 2.4.1 Leave P Out Cross Validation

The practice of withholding a number ($P$) of randomly chosen samples from the data for testing is known as leave $P$ out. The analysis is repeated for all unique combinations of $P$, with the error averaged across trials to give a general error rate. The issue with this approach (i.e., evaluating all combinations) is that, for large values of $P$, the analysis becomes very computationally expensive. A common compromise is the adoption of $P = 1$, otherwise known as Leave-One-Out (LOO) analysis. This is used in study two as compared with larger values of $P$, it provides the maximum amount of training data and limits the number of iterations required to the number of observations. A general disadvantage of the LOO approach is that it still becomes computationally expensive at larger sample sizes, and it only results in a point estimate of accuracy; it is not possible to calculate an interval estimate (e.g., to provide the standard deviation) from the resulting data as each fold represents a binary 'correct'/'incorrect' outcome. As such, the researcher may only provide a single value (e.g., in experiment two the fraction of participants correctly classified) and this does not give insight into how stable the model's performance is across different training and test sets. As such, in larger samples (such as in studies three and four) I chose to apply k-fold cross validation as this technique does provide information about the variability in model performance across different training and test sets.

### 2.4.2 K-Fold Cross Validation

The k-fold cross validation approach refers to the process of splitting the data set into K subsets, termed folds. The algorithm is trained upon all folds except one, and the precision of the resulting model is evaluated upon the remaining fold. In a five fold example, five models formed by training the algorithm upon different but overlapping training folds would be evaluated once each upon their respective (non-overlapping) test fold to yield a measure of out-of-sample performance. This gives five measures of out-of-sample performance, which in this thesis is reported as the mean value with a standard deviation to indicate the variability of performance across different training and test sets.

### 2.4.3 Nested Cross Validation

As discussed in later sections, the algorithms used within this thesis have free parameters ('hyperparameters', examples are given below) which modify how they learn from the provided training data. As the optimal value for each hyperparameter is apriori unknown, it must be estimated from the data. This means that the researcher must estimate the optimal hyperparameters for the algorithm and evaluate the performance of the resultant model upon out-of-sample data. It is essential that the model is not tested upon the same observations that it has been trained upon, as this would no longer represent out-of-sample error. This ultimately reduces the amount of data available to train the model, although a nested cross-validation scheme can be used to minimise the amount of training data sacrificed.

In nested cross validation (illustrated in Figure 2.1), an additional 'inner' cross-validation scheme estimates the optimal hyperparameter values from each training set provided by the outer cross-validation scheme. To demonstrate this, let us imagine a nested 5-fold scheme upon a data set consisting of 100 observations. This means that the algorithm is initially provided 80 observations to learn from, and is tested upon 20 observations. We will call these 80 observations the external training set, and the 20 observations the external test set. The key to nested cross validation is that these 80 observations from the external training set will be further broken into training (N=64) and testing (N=16) sets by an additional cross-validation scheme (five-fold in this example) to estimate the optimal hyperparameter settings. These training and test sets will be denoted as 'internal'. This allows the researcher to try various hyperparameter settings upon the internal training sets and select the parameters that perform best (upon a given evaluation metric) across the internal test sets, without ever including information about the external test set. The algorithm

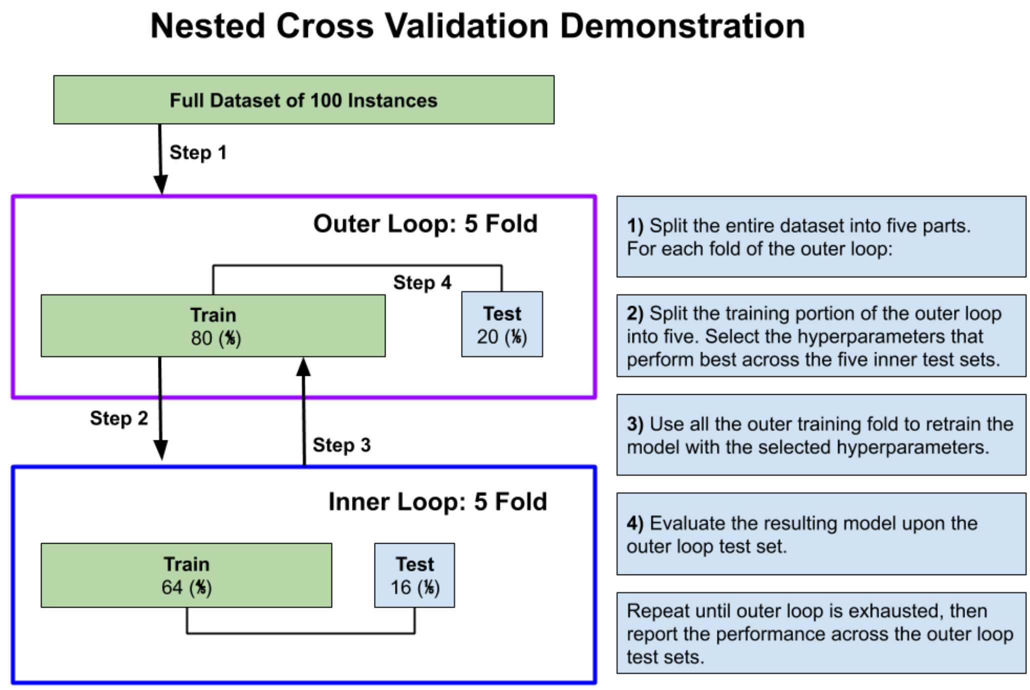


FIGURE 2.1: The nested cross validation procedure

is then trained upon the full (N=80) external training fold using the best performing hyperparameters identified in the previous step. This forms the final model, the performance of which is evaluated upon the external test fold (N=20). Importantly, at no point has the external test fold been used to train the model (e.g., the external test set is still out of sample). This process repeats for every fold of the outer cross validation scheme. As such, the nested cross validation technique is very efficient with how the data is used, making it ideal for maximising the number of observations the algorithm has to learn from.

The main take away point is that cross-validation describes a computational mechanism, the purpose of which is to segregate the entire available data set into discrete parts. The point of cross-validation is to highlight which parts of the data should be provided to the algorithm (i.e., the learning function) to learn from, and which parts of the data the resulting model should be tested upon to measure out-of-sample performance. The efficient use of the available data allows the researcher to highlight the model's performance across many different training and test sets, all whilst incurring a minimal sacrifice to the number of observations available for the algorithm to learn from. This is important as, all held equal, the performance of the machine learning algorithm tends to improve as the number of instances to learn from grows (Alpaydin, 2014; Sun, Shrivastava, Singh, & Gupta, 2017).

### 2.4.4 Machine Learning Algorithms: Classifiers

Having now established the cross-validation scheme used to train and test the models, I will introduce the algorithms used to build these models. Importantly, there is no way to know *apriori* which algorithm will perform best for each of our classification tasks. It may be that one algorithm outperforms all others when predicting one personal attribute from a given set of eye movements, but performs poorly when predicting another personal attribute. This relates to the no free lunch theorem which states that if an algorithm performs well on a certain class of problems, then it necessarily pays for that with degraded performance on the set of all remaining problems (Wolpert, 2002). Accordingly, to be able to capture a variety of associations between the participant's oculomotor behaviour and their attributes (e.g. personality), I chose to apply a variety of algorithms. However, these were not randomly selected but hand-picked, firstly for their ability to learn from relatively few instances, and secondly to represent a wide variety of different methods by which to learn the association between the input metrics (i.e. eye movement descriptors) and the personal attribute being investigated. These algorithms are k-nearest neighbours (Guo, Wang, Bell, Bi, & Greer, 2003), ridge classification (Friedman, Hastie, & Tibshirani, 2010), linear support vector machines (Hearst, Dumais, Osuna, Platt, & Scholkopf, 1998), and naive Bayes (Zhang, 2004). All the above algorithms are readily implemented within scikit-learn, a popular machine learning software package (Pedregosa et al., 2011). Below, I discuss how each algorithm has a different way of learning, and how each complements the other by performing well in cases where the other struggles. As part of this, I will discuss how I searched for the best hyperparameters in each algorithm.

### k-Nearest Neighbours

The k-nearest neighbours algorithm (k-NN) utilises a type of what is termed 'lazy learning', whereby each observation and its associated outcome is memorised (Guo et al., 2003). To make a prediction for a new observation with no known outcome, the algorithm simply looks for the $k$ observations that are nearest to the provided observation (in the feature space defined by the observation's attributes), and predicts that the new observation is the same category as the majority of its neighbours. To avoid ties (i.e., where an equal number of neighbours are associated with different outcomes), it is advisable to choose odd numbers for $k$. As such, in this thesis I explore odd numbers of $k$ between one and twelve $k$ (as $k$ is a hyperparameter). The way which the algorithm determines which neighbour is closest is also a hyperparameter, and in this thesis both Euclidean (shortest path) and Manhattan (rectilinear) distance are considered as ways to determine the closest neighbours. A drawback of k-NN is, due to its need to compute the distances between the new observation and every known observation, it quickly becomes computationally expensive to calculate predictions when applied to very large data sets and at large values of $k$. It also can not identify which features are most important to the prediction (Alpaydin, 2014).

I chose to implement k-nearest neighbours (k-NN) as it is capable of capturing non-linear associations between the input and outcome, and due to our small samples sizes (e.g., less than 200) I do not have to worry about the drawback regarding computational efficiency.

### Ridge Classification

In contrast to k-NN, ridge classification is based upon a regression technique (although, for our purposes the output is still a discrete label), whereby a linear association is sought between the input and output (for a full review, see: Friedman et al., 2010). The advantage of this learning algorithm is that the features can be ranked by importance, which provides additional interpretative power to the researcher. The disadvantage is that this technique is poor at modelling non-linear associations, which makes it an ideal complementary algorithm to apply alongside k-NN. The main hyperparameter of the ridge classification algorithm is the amount of regularisation (i.e., the amount of penalty to apply for large values) to apply to the coefficients, with larger amounts of regularisation pushing coefficients towards zero (and thus forcing the model to be less flexible, which encourages a model that generalises well to out-of-sample instances; Friedman et al., 2010).

**Linear Support Vector Machines**

The linear support vector machine (SVM) algorithm seeks to achieve two objectives. Firstly, to draw a decision boundary in the $n$ dimensional feature space (where $n$ is the number of features) that maximises the minimum distance between the categories (i.e., a maximally separating plane between the points labelled as different categories). Secondly, it seeks to avoid incorrect classifications. Only observations that are proximal to the intersection of the categories within the feature space influence the placement of the decision boundary, which are termed 'support vectors'. The user can choose to prioritise the first objective of a maximally separating plane which seeks a robust solution most likely to generalise to out of sample data. However, this may compromise performance, as a tighter fitting decision boundary may provide fewer mis-classifications within the training data. As such, it is advisable to seek the largest emphasis upon correct classifications that generalises well within the out of sample test data. The tuning parameter for this trade-off is termed $C$, whereby low values of $C$ emphasise the first objective and high values emphasise the second objective (Pedregosa et al., 2011).

The linear support vector machine algorithm is less susceptible to outliers than the above mentioned ridge classification algorithm (which minimises the overall error across all observations). As such, this algorithm complements the ridge classification algorithm by performing well within cases where it is best to concentrate upon only observations proximal to the boundary between categories rather than to all samples (i.e., to ignore large outliers). For a full review of linear support vector machines, see Hearst et al. (1998).

**Naive Bayes**

Naive Bayes calculates the likelihood of observing a particular value for each feature, given that it belongs to a category. This allows it to evaluate the evidence that an observation belongs to each category and predict the category with the highest likelihood. To achieve this, it assumes that features do not interact (i.e., conditional independence) and that the likelihood follows a user-specified distribution. In this thesis, it is assumed that the likelihood is normally distributed. Despite conditional independence being rare in real world data sets (and is violated in this research, as fixations and saccades are related events), the algorithm has been found to perform well in various classification tasks where the (naive) assumption of conditional independence is violated (e.g., medical diagnosis tasks; Rish, 2001; Zhang, 2004). As such, this technique is still applicable and complements the previous algorithms as it is driven by probabilistic evidence rather than trying to find linear decision boundaries (e.g., as in ridge and support vector machine algorithms) or employing a 'lazy learning' technique (e.g., as in k-NN).

### 2.4.5 Preprocessing

Before applying the algorithms listed above, it is important to first process the collected eye data into a machine-friendly format. A classifier built using any of the above algorithms would not be able to provide a prediction for an observation with missing values (e.g., in any of the eye movement descriptions), as the observation would not be defined within the same space as the classifier's decision boundary. A common solution to missing values is to either remove the observation or impute (i.e., fill in) the value (Alpaydin, 2014). It is also important to scale features such that they are expressed in the same units, else the distance between observations can become dominated by the units with a large scale. This leads algorithms such as support vector machines and k-nearest neighbours to require features to be scaled for optimal performance (Alpaydin, 2014). A common method is to scale features such that they have a mean of zero and one unit standard deviation, termed 'standard scaling' or standardisation within the machine learning literature (Alpaydin, 2014). In this thesis, this is how I standardise the eye movement features before applying any machine learning algorithm.

Having selected the algorithms that will learn from the data and how to prepare the eye movement data into a machine-friendly format, all that remains is to evaluate the model's performance. Evaluating performance is important for understanding the final classifier (i.e., when and how often is it right?), but also for selecting the hyperparameters (by definition, the 'best performing') that are used to form the final classifier. As such, in the next section I discuss the different evaluation metrics considered in this thesis.

### 2.4.6 Model Evaluation: Metrics and Methods

Accuracy, defined as the percentage of correct predictions made, provides an intuitive summary of model performance by answering the question 'how often is the model right?'. Whilst intuitive and straightforward to interpret, accuracy does have shortcomings as an evaluation metric. It does not give any insight into when the classifier gets a prediction wrong and may be misleading in situations where one category is substantially more frequent than the others. For instance, a classifier may achieve a reasonably high accuracy by always predicting the most frequently occurring category, which results in a 100% error rate for the less frequent category(s). As such, when reporting accuracy metrics in this thesis, I also report the proportion of samples in the most frequent category. Furthermore, to provide insight into when the classifier makes incorrect predictions, in study two I also report a metric called the receiver operator characteristic, which is explained below.

#### Receiver Operator Characteristics and F1

When a binary classifier provides an incorrect prediction, it may be either a false positive (e.g., the participant is predicted as being high upon trait openness, but is actually low) or a false negative (e.g., predicted low, but is high). The researcher can modify the threshold at which the classifier will predict an outcome as being positive (e.g., requiring more/less evidence for a positive prediction), which influences the trade off between capturing as many positive cases as possible ('recall'), and minimising the number of false positives ('precision'). The 'best' value depends upon whether the user values recall ($\frac{TruePositives}{TruePositive+FalseNegatives}$) over precision ($\frac{TruePositives}{TruePositives+FalsePositives}$). The receiver operator characteristic (ROC) is the trade off between true and false positives for a range of different threshold values, and the area under the ROC curve (AUROC) provides an insight into the overall performance of the model across all decision threshold values. For a review of ROC methodology, analysis and examples of applications, see Hajian-Tilaki (2013) and Huang and Ling (2005). A drawback to the ROC approach is the focus upon evaluating binary choice decisions, for which it is most suited. As such, in study two (a binary classification task) I report the AUROC metric (reflecting the general performance of the model across all decision thresholds). However, in studies three and four, I have more than two categories, which makes the AUROC difficult to interpret. An alternative is the $F1$ metric, calculated as the harmonic mean of precision and recall. Importantly, the $F1$ score can be meaningfully generalised to the multi-category case by calculating the $F1$ score for each category, and then taking the average of these values (termed the $F1_{macro}$ score). As such, I report the $F1_{macro}$ metric for studies three and four.

Having established the evaluation metrics, it is important to understand whether the score upon the metric is significantly better than would be expected from chance. As the machine learning framework is non-parametric this requires permutation based methods (described in the following section) to be used.

**Label Permutation Test**

When we permute the label assignments between observations, we break the observed association between the outcomes and the associated feature values. By repeatedly permuting the labels across observations and recording the number of instances where the randomly assigned class is the same as the true class, we establish the distribution of the null hypothesis - the spread of classification scores obtained under the assumption that there is no real connection between the data and class labels (Ojala & Garriga, 2010). We can then compare the performance of our classifier to this null distribution, and identify the probability of obtaining the given model's performance within the data provided. Our hypothesis is that the model has found a real connection between the data and class labels, and we reject the null if our p-value exceeds our alpha value (Equation 2.1, simplified from Ojala & Garriga, 2010). This provides a method of bridging the gap between conventional hypothesis testing and machine learning studies, and is used extensively in this thesis (with 200,000 permutations) to quantify if personal attributes can be predicted significantly better than chance from oculomotor behaviour.

$$p = \frac{(\sum_{i=1}^{N}(PermutationScore_i \geq ModelScore)) + 1}{N + 1} \tag{2.1}$$

## 2.5 Study One: Does emotional valence predict oculomotor behaviour whilst browsing social media?

The first question I address in this thesis is whether the same link between visual behaviour and preference decisions (e.g., as found in Holmes & Zanker, 2012; Maughan et al., 2007; Nummenmaa et al., 2006) emerges whilst participants browse a novel SNS visual environment. As discuss earlier, three visual metrics were explored (time to first fixation, total fixation duration, number of fixations). A key difference between static image design and the SNS visual environment is the ability to scroll to currently occluded content. As such, to answer the question, a critical part of this study was to design a stimulus which the participant could explore by scrolling. To achieve this, I developed a mock SNS webpage for the participant to browse through.

### 2.5.1 Stimulus Presentation: Designing a mock social media webpage

To allow the participant to scroll through the content whilst maintaining experimental control (i.e., to prevent content changing across participants), I created a simplified offline webpage by hosting a large image (manually created by stitching together social media posts) within a HyperText Markup Language (HTML) table. This substantially reduced load times and increased responsiveness compared to simply downloading the entire webpage, mostly by reducing the CPU load associated with dynamic HTML elements. This was important, as it prevented delays in updating the visual scene when scrolling. The compromise was relatively small in terms of ecological validity, for example, the side bars no longer track along the page as the participant scrolled. During piloting, I asked participants to list any differences they noticed between this webpage and their own social media feed, and discovered that participants did not notice the difference in tracking side bars.

### 2.5.2 Counterbalancing to avoid order effects

In contrast to the sequential or simultaneous presentation of static images used within previous literature (e.g., Holmes & Zanker, 2012; Nummenmaa et al., 2006), in my social media themed webpage the participants were able to scroll and discover new content at their leisure. This influences the metric of time to first fixation, which in our paradigm is likely to simply reflect the content's position upon the webpage rather than the participant's preference decision (e.g., as shown in: Holmes & Zanker, 2012; Nummenmaa et al., 2006). As such, the location of the stimulus provides a substantial confound when interpreting a significant result, as we can not disentangle whether the association has emerged due to the order of contents upon the page or due to the affective response of the participant being reflected in their oculomotor behaviour. Furthermore, the number of fixations and the amount of time spent upon a stimulus may also be influenced by page position relative to the initial fixation, as the participant may have less opportunity to view later content within their thirty second browsing duration. To address this, in a repeated presentation design, participants were shown the social media themed webpage five times, each time viewing a rotated order of the content upon the page. This resulted in the creation of five different social media style webpages. The presentation sequence is randomly allocated across participants to avoid order effects. In this way, each of the five social media posts has an equal opportunity to be seen first across the presentations. Importantly, I calculate the three oculomotor metrics (see above) across the five presentations (taking the mean value), which results in any remaining association between the visual metrics of interest and subjective rating being dissociated from the content's location upon the page.

### 2.5.3 Visualisation: Gaze Opacity Map

The counterbalancing procedure outlined above should result in similar viewing patterns across each of the five webpages when visual behaviour is aggregated across participants. Importantly, there should be no clear biases in the distribution of fixations, where one version of the webpage is viewed differently from the rest. To ensure data quality, I calculated a gaze opacity map for each webpage, as this allowed me to visually inspect all five webpages and confirm that similar overall viewing patterns occurred on each page.

A gaze opacity map is a visualisation technique where an individual or a cohort's fixations (in this study, fixation duration summed across the cohort) are mapped to the pixels within the scene as values in a 2D matrix, and then smoothed with a Gaussian filter. This leads to peaks where the participants have looked the most, which decay away according to the smoothing kernel. In gaze opacity maps, a black image is overlaid upon the reference image, and as the density of the fixation duration increases, the opacity of the matching location in the black overlay reduces (becomes more transparent). This allows content in the reference image to 'pop through' the black overlay, which highlights where the individual's have attended. As fixation behaviour is often seen as a correlate of attention (Hayhoe & Ballard, 2005), it seems fitting that regions that do not receive visual attention appear opaque and unknown. As such, the gaze opacity map effectively highlights regions that have, and have not often been attended to within the novel social media style webpage.

### 2.5.4 Analysis Techniques

Having created the stimuli and visually inspected the eye tracking data, I was able to continue with the analysis. In this first experiment, I began by using multiple linear regression. I employed multiple linear regression as it allowed for a fully interpretable model, where a unit increase in each predictor (e.g., fixation duration) could be associated with a unit change in the outcome variable. Furthermore, using this technique, I can establish both how likely it would be to find the overall model fit by chance and whether each metric individually accounts for a significant amount of variance in the outcome variable. This uniquely allows for a direct comparison to previous literature which has followed a similar procedure (e.g., Holmes & Zanker, 2012; Maughan et al., 2007; Nummenmaa et al., 2006).

Linear regression makes strong assumptions about the underlying data being modelled. Namely, it assumes that the association between the outcome and predictors is linear in nature, requiring that the model residuals are normally distributed (multivariate normality), that the predictors are not highly correlated with each other (no multicollinearity), and that the model is roughly equally competent at predicting the outcome across the range of predictor variables values (i.e., meets the assumption of homoscedasticity). The assertion that a linear association exists between the eye movement metrics described and emotional valence rating is supported by previous literature, with a wide range of studies finding a significant linear association between the aforementioned metrics and subjective ratings (e.g., Holmes & Zanker, 2012; Maughan et al., 2007; Nummenmaa et al., 2006). In this study, I identified that the outlined assumptions were met, which allows for the method to be applied.

## 2.6 Study Two: Eye movements on a mock SNS page predict personality traits

Using the same data set as in study one, in this study, I created a broad set of eye movement metrics (in machine learning terms 'features') and used them to predict whether the individual was either low or high for each of the five personality traits. As such, this study seeks to answer whether aspects of an individual's personality can be predicted from their browsing behaviour upon SNS type content.

### 2.6.1 Describing eye movements: AOI Metrics

To begin with, building upon the previous study, I described the time to first fixation (TTFF), total fixation duration (TFD), and number of visits associated with each social media post. The number of visits metric was selected as it represented the number of times the participant viewed the content and it was available in the Tobii Studio software used to calculate these metrics. Notably, this metric replaces the number of fixations metric used in the previous study. There are five social media posts, and each has an interaction section underneath which shows how other individuals have reacted to the post (e.g., via commenting or liking). As some of the outcomes are regarding how individuals seek social information (i.e., whether they are more or less extroverted), I also calculate the TTFF, TFD, and the number of visits associated with each interaction section. This resulted in ten areas of interest (AOI), each described by three metrics, for each of the five displayed webpages (totalling 150 features). I then averaged each metric across the five presentations following the steps outlined in the previous study (resulting in 30 features). I then repeated this procedure to calculate the same metrics, but instead encoded by the AOI's location (i.e., top, middle, bottom, etc.) rather than its content. To compare the performance of classifiers trained upon the content-based descriptions with those trained upon location-based descriptions, I separate the features into two discrete groups ('Content-Encoded' and 'Location-Encoded' respectively, each containing 30 features).

### 2.6.2 Describing eye movements: Statistical Metrics

I created a third set of features that provide statistical (minimum, maximum, count, 25th percentile, 50th percentile, 75th percentile, mean, and standard deviation) descriptions of the fixation and saccadic properties. Before using these metrics to train the machine learning models, I removed the feature of minimum saccade duration as it was identical across participants (being defined by the fixation filter) and, thus, offered no discriminative power. Similarly, I also removed the feature of the number of saccadic amplitudes, as it is perfectly correlated with the number of fixations (the amplitude is calculated as the distance between two fixations). After these subtractions, a total of 22 metrics remained in the statistical feature group.

A particular advantage of testing separate sets of features is that we can infer the relative insight provided by each set separately. As such, by employing this approach, we gain additional insight into which descriptions of oculomotor behaviour best predict each personality trait. This can be compared to the way in which a multiple linear regression provides the separate contribution of each predictor, except now I compare sets of features instead. Finally, in an exploratory analysis, I combine the feature groups together to understand whether this improves predictive power.

### 2.6.3 Model Evaluation

To evaluate the performance of the classification models, I used a nested cross-validation approach. I choose to utilise a LOO outer scheme to maximise the number of training instances, and a 5-fold inner scheme for estimating the optimal hyperparameters (evaluation metric: accuracy). In this way, the algorithm is provided with 34 instances to train upon (providing 20% more data than, for example, a five-fold scheme), and one instance to test upon in the outer fold. To aid interpretation, in this paper I report both the classifier's accuracy and the accuracy achieved by the most frequent class strategy. I calculate the AUROC score if the classifier's accuracy is significantly better than chance (as indicated by the label permutation test).

## 2.7 Study Three: Twenty Seconds to Know You? Classifying Personality from Visual Behaviour on Social Media

In previous studies, I investigated whether personality trait information can be decoded from the participant's eye movements whilst they browse a mock social media themed webpage. The mock webpage is not truly ecologically valid, as compromises were made for performance and to provide a measure of experimental control. Taking the lessons learned (i.e., machine learning algorithms, description techniques, and evaluation metrics) from previous studies, this study extends my research into the ecologically valid setting of the participant's own Facebook page in a free viewing design. This means that the type of content is not experimentally controlled and varies across participants. As this provides a more difficult classification task, I collected a relatively large sample size to provide more observations for the machine learning algorithms to learn from.

### 2.7.1 Data Preparation

Due to the encouraging results of the previous paper (study two), a more challenging three-class (low, medium, high) classification task (e.g., as selected in; Berkovsky et al., 2019; Hoppe et al., 2018) was chosen to provide greater insight into the individual's personal attributes. To achieve this three-category split from the continuous personality trait scores, I again applied a quantile-based technique to equally (to the extent allowed by discrete values) assign the observations into the low, medium and high categories and check if the groups are significantly different.

**Feature Engineering: AOI Metrics**

To enable content-based descriptions of visual behaviour (as employed within previous studies), the social media content seen by the participant was manually labelled as belonging to one of the seven content categories: 1) create posts; 2) image; 3) text; 4) video; 5) hybrid (i.e., an image overlaid with text); 6) comments; 7) interaction elements (see Figure 5.1 in chapter 5). I calculated the time to first fixation, total fixation duration (TFD) and number of fixations in relation to each content category. The number of fixations metric was selected (and replaces the number of visits metric used earlier) as we conducted our own preprocessing (via custom python scripts) and decided that it was the more commonly used and interpretable metric.

Of note is that the content categories do not all occupy the same amount of the page (e.g., interaction elements have a smaller footprint than an image), and for some participants some categories may not occur at all. As this creates missing values, to convert the data into a machine friendly format, I imputed the TFD and the number of fixation metrics with zero when the content category did not occur, and the TTFF metric with the maximum presentation duration of 20 seconds. This creates the 'AOI' feature group, consisting of 21 features. However, these metrics do not contain information about how often each category occurred, or regarding the amount of the page occupied by the given category. To reflect how much of the page each content type accounted for, I created a new feature group where each metric (from the AOI feature group described above) is reweighted (i.e., multiplied) by the proportion of the page occupied by the content type (individually for each participant). I named this group 'AOI Proportional', and it contained 21 features.

Similarly, to provide information about how often each content type occurred, I calculated the number of times each category was encountered for each participant. Since there are seven content categories, this creates seven new features. I created two new feature groups by combining these seven metrics with the AOI feature group (forming the feature group 'AOI with Frequency', 28 features), and the AOI Proportional feature group (forming the feature group 'AOI Proportional with Frequency', 28 features). This results in a total of four AOI-based feature groups, each containing different information about how the participant reacted visually to the content upon the page. Finally, to understand the insight provided by knowing the page content alone (and not the visual behaviour) I included a control feature group which consisted of the proportion of the page occupied, and the frequency of occurrence, of each content category. By comparing the performance of classifiers trained upon these different feature groups, I am able to gain a deeper insight into which characterisations of visual behaviour best reflect a given trait.

#### Feature Engineering: Statistical Features

As in my previous studies, I created a set of statistical features. For each property of fixations and saccades, I calculated the sum, count, minimum, maximum, mean, standard deviation, range, and interquartile range. I also include the mean progress in vertical screen-based coordinates per second (pixels per second) across the viewing duration, as this provides an index of how rapidly the participant progressed through the webpage. As before, I removed the features of minimum saccade duration and number of saccadic amplitudes. This results in a total of 22 features in the statistical feature group. Finally, to understand the insight provided by knowing the page content alone (and not the visual behaviour), I included a control feature group which consisted of the proportion of the page occupied and the frequency of occurrence, of each content category (14 features).

#### Model Evaluation

The same machine learning algorithms were applied as in study two (see: 2.4.4), but a different model evaluation scheme was employed. Whilst k-fold cross-validation was unsuitable for the outer loop in the previous study (as it left too few instances to learn from), in this study a five-fold approach yielded 135 training instances, and 45 testing instances. In our case, this provides a reasonable amount of data to both train (including estimation of optimal hyperparameters) and evaluate the model upon when employing a five-fold nested cross validation procedure (outer: five-fold; inner: five-fold). The advantage of this is that is provides an interval based estimate, which provides the researcher with an insight into how stable the model's performance was across different training and testing sets. A secondary additional benefit is that it allowed me to optimise the algorithm not upon accuracy but upon the $F1_{macro}$ metric. This ensured that the algorithms were encouraged to learn decision boundaries that were equally good at predicting each category, rather than focusing upon the majority case. Finally, a practical benefit is that this approach is also much more computationally efficient, which allows the experiment to be replicated on an ordinary desktop PC in a reasonable time frame. Of note is that I did not report AUROC metrics in this experiment. This is because such results are not as easily interpreted in the three category case, which is a drawback that comes from searching for a more fine grained insight into the participant's personality attributes.

## 2.8 Study Four: The effect of labelling strategy upon predicting private attributes from visual behaviour

In previous literature (including the two machine learning studies reported in this thesis), authors have often split the participant's personality trait scores into discrete categories (e.g., [low, high] or [low, medium, high]). This is often achieved via a quantile based splitting strategy (e.g., as in Berkovsky et al., 2019; Hoppe et al., 2018), but little rationale is given behind why the authors choose this strategy or how they choose the number of categories. In this study, we investigate whether the application of machine learning techniques can provide a data-driven justification for choosing one strategy over the other and for choosing the number of categories. To this end, I apply two unsupervised (i.e., learning based upon the features alone, with no known output category) machine learning algorithms to assign our participants into categories depending upon their personality trait score (this task is termed 'clustering'). I then investigate whether this more rigorous method of assigning participants to personality trait categories influences the performance of the classifiers when predicting the given outcomes from visual behaviour. Finally, using the lessons learnt from the previous analysis, I investigate whether the attributes of sex, self-esteem, and voting behaviour (collected from participants in the previous study) can be predicted from visual behaviour upon a social media webpage. Exploring these additional attributes extends the scope of the project beyond personality prediction, with each of the chosen attributes having theoretical links to how and why an individual engages in SNS use (outlined in detail within the paper). The same algorithms used in studies two and three are employed here. The feature groups from study three ('Eye Movement Statistics', 'AOI', 'AOI Proportional', 'AOI with Frequency', 'AOI Proportional with Frequency', 'Control Group') were used here also.

The algorithms used to perform the unsupervised clustering task (i.e., assigning participants into groups based upon their features alone) are k-means and agglomerative clustering. Chapter 6 itself provides a great amount of detail regarding the application of the following methods, thus only a brief description of each algorithm is provided here to explain the methods followed.

**K-Means Clustering**

The k-Means algorithm seeks to assign a set of observations into $k$ categories (where $k$ is predefined by the researcher) in such a way as to minimise the distance between observations within the same category. In the machine learning literature, these are termed clusters. To provide a simplified overview, the algorithm achieves this by randomly assigning $k$ points ('centroids') within the feature space defined by the observations' attributes. Observations are assigned to the centroid closest to them. The algorithm then repeatedly moves the centroids until it can no longer reduce the within-category sum of squares. In this study, the participants are grouped upon a single one-dimensional attribute (e.g., a questionnaire score).

**Agglomerative Clustering**

The agglomerative clustering algorithm implemented within this thesis works upon a simple principle; it first assigns every observation to its own cluster, then successively merges nearby clusters (i.e., in a bottom-up manner, see Figure 6.4 in Chapter 6). The way that clusters are merged depends upon how the distance between the clusters is measured, called the linkage function. Many functions are available, but perhaps the most common are the single-linkage, complete linkage, and ward linkage functions (Alpaydin, 2014). The single linkage function measures the smallest distance between two observations in different clusters (e.g., neighbouring clusters are joined). Complete linkage function measures the largest distance between two clusters. Finally, the ward linkage function considers the impact that merging two clusters would have upon the sum of the squared distance between all observations in that cluster (Ward, 1963). As the objective of this study is to find clusters that are internally coherent, I utilised the ward linkage function.

The agglomerative algorithm is also known as hierarchical clustering, due to the hierarchical structure produced by each successive join (i.e., the merging of two clusters). This algorithm has two key attributes relevant to this thesis. Firstly, in contrast to k-means, the algorithm does not require the researcher to predefine the number of clusters to form. Secondly, due to its sequential nature, the final solution is constrained by earlier joins. To elaborate, once an observation is joined to a cluster, it can not be reassigned to a different one at a later stage even if this improves the final solution.

In this chapter, I utilise the agglomerative clustering algorithm to understand the structure inherent within the data, which informs the number of categories to form. Finally, I apply the k-means algorithm to find a good solution for the given value of $k$ (as the k-means algorithm is not constrained by previous merge decisions). This is a particularly novel application of these techniques within the literature and has some inherent assumptions that may preclude application in the more general case. For instance, the k-means algorithm operates within Euclidean space and assumes that the true cluster structure radiates away from the center point equally in all directions (i.e., it assumes the cluster is isotropic). Similarly, the ward linkage function also shares this assumption (Ward, 1963). This can cause problems when the true cluster structure is not isotropic (e.g., an elongated sphere) or within a very high-dimensional space where Euclidean distance may be inflated (Alpaydin, 2014). However, since we are clustering in one-dimensional space in this study (i.e., upon the final scores for questionnaire based outcomes), neither the isotropic assumption or inflation of Euclidean distance is of concern. As such, these techniques are suitable for application within this study, but may not always generalise well to applications within multiple dimensions.

## 2.9 Conclusion

I have now laid out the rationale and background details regarding the eye tracking techniques applied (including the hardware, software, and filter algorithm choices), the stimuli presented to the participants, and how I have described the visual behaviour expressed by participants. I have also introduced why and how machine learning has been employed within this thesis, including the algorithms used, cross-validation techniques applied, and the evaluation metrics selected. This provides the reader with the necessary information to follow the next sections, which contain the four independent studies reported within this thesis.

## References

Alpaydin, E. (2014). *Introduction to machine learning*. MIT press.

Ames, D. R., Rose, P., & Anderson, C. P. (2006). The NPI-16 as a short measure of narcissism. *J. Res. Pers. 40*(4), 440–450. doi:10.1016/j.jrp.2005.03.002

Anderson, B. (2013). A value-driven mechanism of attentional selection. *J. Vis. 13*(3), 7–7. doi:10.1167/13.3.7

Andersson, R., Larsson, L., Holmqvist, K., Stridh, M., & Nyström, M. (2017). One algorithm to rule them all? An evaluation and discussion of ten eye movement event-detection algorithms. *Behav. Res. Methods*, *49*(2), 616–637. doi:10.3758/s13428-016-0738-9

Berkovsky, S., Taib, R., Koprinska, I., Wang, E., Zeng, Y., Li, J., & Kleitman, S. (2019). Detecting Personality Traits Using Eye-Tracking Data. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (pp. 1–12). CHI '19. doi:10.1145/3290605.3300451

Costa, P. T. (1996). Work and personality: Use of the NEO-PI-R in industrial/organisational psychology. *Appl. Psychol. 45*(3), 225–241. doi:10.1111/j.1464-0597.1996.tb00766.x

Duchowski, A. T. (2017). Taxonomy and Models of Eye Movements. In A. T. Duchowski (Ed.), *Eye Tracking Methodology: Theory and Practice* (pp. 39–46). doi:10.1007/978-3-319-57883-5_4

Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of statistical software*, *33*(1), 1–22.

Gosling, S. D., Rentfrow, P. J., & Swann, W. B. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in Personality*, *37*(6), 504–528. doi:10.1016/S0092-6566(03)00046-1

Guo, G., Wang, H., Bell, D., Bi, Y., & Greer, K. (2003). KNN Model-Based Approach in Classification. In *On the Move to Meaningful Internet Systems* (pp. 986–996). doi:10.1007/978-3-540-39964-3_62

Hajian-Tilaki, K. (2013). Receiver Operating Characteristic (ROC) Curve Analysis for Medical Diagnostic Test Evaluation. *Casp. J. Intern. Med. 4*(2), 627–35.

Hayhoe, M., & Ballard, D. (2005). Eye movements in natural behavior. *Trends Cogn. Sci. 9*(4), 188–194. doi:10.1016/j.tics.2005.02.009

Hearst, M. A., Dumais, S. T., Osuna, E., Platt, J., & Scholkopf, B. (1998). Support vector machines. *IEEE Intelligent Systems and their Applications*, *13*(4), 18–28. doi:10.1109/5254.708428

Hessels, R. S., Kemner, C., van den Boomen, C., & Hooge, I. T. (2016). The area-of-interest problem in eyetracking research: A noise-robust solution for face and sparse stimuli. *Behav. Res. Methods*, *48*(4), 1694–1712. doi:10.3758/s13428-015-0676-y

Holmes, T., & Zanker, J. M. (2012). Using an Oculomotor Signature as an Indicator of Aesthetic Preference. *i-Perception*, *3*(7), 426–439. doi:10.1068/i0448aap

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & van de Weijer, J. (2011). *Eye Tracking: A Comprehensive Guide To Methods And Measures*.

Holmqvist, K., Nyström, M., & Mulvey, F. (2012). Eye tracker data quality. In *Proc. Symp. Eye Track. Res. Appl. - ETRA '12* (p. 45). doi:10.1145/2168556.2168563

Hoppe, S., Loetscher, T., Morey, S. A., & Bulling, A. (2018). Eye Movements During Everyday Behavior Predict Personality Traits. *Front. Hum. Neurosci. 12*, 105. doi:10.3389/FNHUM.2018.00105

Huang, J., & Ling, C. (2005). Using AUC and accuracy in evaluating learning algorithms. *IEEE Transactions on Knowledge and Data Engineering*, *17*(3), 299–310. doi:10.1109/TKDE.2005.50

Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nat. Rev. Neurosci. 2*(3), 194–203. doi:10.1038/35058500

Maughan, L., Gutnikov, S., & Stevens, R. (2007). Like more, look more. Look more, like more: The evidence from eye-tracking. *J. Brand Manag. 14*(4), 335–342. doi:10.1057/palgrave.bm.2550074

McCrae, R., & Costa, P. (2004). A contemplated revision of the NEO Five-Factor Inventory. *Pers. Individ. Dif. 36*(3), 587–596. doi:10.1016/S0191-8869(03)00118-1

Nummenmaa, L., Hyönä, J., & Calvo, M. G. (2006). Eye movement assessment of selective attentional capture by emotional pictures. *Emotion*, *6*(2), 257–268. doi:10.1037/1528-3542.6.2.257

Ojala, M., & Garriga, G. C. (2010). Permutation Tests for Studying Classifier Performance. *J. Mach. Learn. Res. 11*(Jun), 1833–1863.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res. 12*(Oct), 2825–2830.

Raeder, T., Hoens, R., & Chawla, N. V. (2010). Consequences of Variability in Classifier Performance Estimates. In *2010 IEEE Int. Conf. Data Min.* (pp. 421–430). doi:10.1109/ICDM.2010.110

Rauthmann, J. F., Seubert, C. T., Sachse, P., & Furtner, M. R. (2012). Eyes as windows to the soul: Gazing behavior is related to personality. *J. Res. Pers. 46*(2), 147–156. doi:10.1016/j.jrp.2011.12.010

Rello, L., & Ballesteros, M. (2015). Detecting readers with dyslexia using machine learning with eye tracking measures. In *Proc. 12th Web All Conf. - W4A '15* (pp. 1–8). doi:10.1145/2745555.2746644

Rish, I. (2001). An empirical study of the naive Bayes classifier. In *JCAI 2001 workshop on empirical methods in artificial intelligence* (Vol. 3, pp. 41–46). IJCAI.

Robins, R. W., Hendin, H. M., & Trzesniewski, K. H. (2001). Measuring global self-esteem: Construct validation of a single-item measure and the Rosenberg Self-Esteem Scale. *Personal. Soc. Psychol. Bull. 27*(2), 151–161. doi:10.1177/0146167201272002

Stuart, S., Hickey, A., Vitorio, R., Welman, K., Foo, S., Keen, D., & Godfrey, A. (2019). Eye-tracker algorithms to detect saccades during static and dynamic tasks: A structured review. *Physiological Measurement*, *40*(2), 02TR01. doi:10.1088/1361-6579/ab02ab

Sun, C., Shrivastava, A., Singh, S., & Gupta, A. (2017). Revisiting Unreasonable Effectiveness of Data in Deep Learning Era. *Proc. IEEE Int. Conf. Comput. Vis. 2017-Octob*, 843–852. doi:10.1109/ICCV.2017.97

Tatler, B. W., Wade, N. J., Kwan, H., Findlay, J. M., & Velichkovsky, B. M. (2010). Yarbus, eye movements, and vision. *Iperception. 1*(1), 7–27. doi:10.1068/i0382

Tseng, P.-H., Cameron, I. G. M., Pari, G., Reynolds, J. N., Munoz, D. P., & Itti, L. (2013). High-throughput classification of clinical populations from natural viewing eye movements. *J. Neurol. 260*(1), 275–284. doi:10.1007/s00415-012-6631-2

Ward, J. H. (1963). Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association*, *58*(301), 236–244. doi:10.1080/01621459.1963.10500845

Wolpert, D. H. (2002). The Supervised Learning No-Free-Lunch Theorems. In R. Roy, M. Köppen, S. Ovaska, T. Furuhashi, & F. Hoffmann (Eds.), *Soft Computing and Industry: Recent Applications* (pp. 25–42). doi:10.1007/978-1-4471-0123-9_3

Yarkoni, T., & Westfall, J. (2017). Choosing Prediction Over Explanation in Psychology: Lessons From Machine Learning. *Perspect. Psychol. Sci. 12*(6), 1100–1122. doi:10.1177/1745691617693393

Zhang, H. (2004). The Optimality of Naive Bayes. *AAI*, *1*(2), 6.

Zhang, Y., & Yang, Y. (2015). Cross-validation for selecting a model selection procedure. *J. Econom. 187*(1), 95–112. doi:10.1016/J.JECONOM.2015.02.006

# Chapter 3

# Does emotional valence predict oculomotor behaviour whilst browsing social media?

## 3.1 Abstract

Where we attend in a visual scene is influenced by our personal preferences, social cues, information requirements, and the physical properties of the scene. In a free-viewing design, we asked 38 participants to scroll through (allowing both scrolling up and down) a mock social media page with five posts while tracking their eye movements. We visualise this viewing behaviour and then explore the association between visual behaviour (time to first fixation, total fixation duration, number of fixations) and preference by asking participants to report their affective response to each post. We identify that preference decisions can be predicted significantly better than chance as a linear combination of these visual metrics, with time to first fixation being the most useful predictor of preference. We find our results to be distinct from previous literature which found fixation duration to be a significant predictor of valence when appraising static images, and discuss how the unique constraints induced by the visual environment may contribute to this result. Finally, we present clear guidelines for future research wishing to predict preference decisions from visual behaviour while participants scroll through social media style content.

## 3.2 Introduction

Where we look in a visual scene can be considered an index of attention, with the average human sampling from the surrounding visual environment at a rate of roughly two to three fixations a second (Kowler, 2011). The expression of visual behaviour is influenced by many factors, such as the cognitive task (Tatler, Wade, Kwan, Findlay, & Velichkovsky, 2010), the physical properties of the scene (Itti & Koch, 2001) and what we find rewarding within the scene (Anderson, 2013). Over time, we become more efficient in our visual search patterns by learning from rewarding stimuli in a given environment (Anderson, 2013; Niu, Todd, Kyan, & Anderson, 2012), which allows visual attention to be influenced by what the individual has found previously rewarding in the visual scene (Anderson, 2013; Chelazzi, Perlato, Santandrea, & Della Libera, 2013; Kelley & Yantis, 2010). For example, learning that a particular type of stimulus leads to rewarding outcomes improves the efficiency of its selection from amongst competing items, and makes it more distracting when not task-relevant (della Libera, Perlato, & Chelazzi, 2011). It is this reward-mediated mechanism of attentional capture which which underpins a common assumption in eye tracking for market research, namely, that individual's tend to preferentially allocate their visual attention to objects or locations which they prefer. This is because it illustrates a mechanism by which oculomotor behaviour may reflect and provide insight into the cognitive (including emotional) associations an individual holds with visual stimuli (Niu et al., 2012).

The assumption that individuals tend to preferentially allocate their visual attention to objects or locations which they prefer is supported by a wide range of literature based upon the viewing of static images. For example, in a study investigating the correlation between fixation duration and subjective evaluation of poster advertisements upon bus shelters, (Maughan, Gutnikov, & Stevens, 2007), participants viewed a series of still images shot from the perspective of a passing car. The authors found a positive correlation between the evaluation of the advertisement and viewing behaviour, such as that positive evaluations were associated with more frequent and longer fixations upon the advert. A similar result for multiple simultaneously displayed stimuli was found, where, independent of whether two, four, or eight images were presented radially around a central fixation cross, participants spent significantly more time fixating images that they subsequently rated as being more attractive (Holmes & Zanker, 2012). This converging evidence points towards a positive correlation between the individual's affective evaluation of a visual location and the number (and duration) of fixations exhibited at the location.

It may be that individuals demonstrate preferential oculomotor behaviour towards viewing emotional content, regardless of whether the emotional salience is positive or negative. Nummenmaa, Hyönä, and Calvo (2006) measured the eye movements of participants as they viewed unpleasant, neutral, or pleasant pictures. They employed a series of trials, each time presenting two images simultaneously upon the diagonal either side of a central fixation cross, one image being always neutral and the other either unpleasant or pleasant, while controlling for low level saliency factors such as luminance and contrast. They found that participants were more likely to first attend to, and make more fixations upon, pictures depicting emotional content (regardless of being positive or negative) compared to emotionally neutral pictures.

In this study, we explore whether visual behaviour (total fixation duration, time to first fixation, number of fixations) is related to affective judgement responses when the stimuli are presented within a novel real-world setting; a scrolling social media web page containing five centrally presented social media posts. This is important, as social media content has many attributes which are novel from the static images shown within previous literature. In contrast to previous literature where images are either sequentially or simultaneously presented, the scrolling design of a social media style web page (with content being occluded until scrolled to) allows the participant to decide when to receive new visual content, rather than the presentation being preordained by the researcher. As such, social media web pages provide a unique medium for exploring the complex interaction between eye movements and our affective judgements in a real-world setting.

This novel visual environment is important as Tatler, Hayhoe, Land, and Ballard (2011) propose that the lack of a dynamic component may contribute towards explaining why studies investigating gaze allocation using static pictures sometimes fail to replicate when investigated in ecologically valid, non-laboratory settings.

Namely, the authors highlight that eye movements within ecologically valid environments are often influenced by the dynamically unfolding nature of the task. For example, when making a drink the individual may look at a cup when they are about to pour water into it, then at the tap before turning the water off. In these naturalistic situations the eye movements guide actions within a dynamically changing environment. Similarly, in our paradigm the participant may first look at a given piece of content (i.e., 'post'), then decide whether to stop and investigate further, or move on to discover a new (currently unknown) post, or return to a previously viewed post. As such, our study provides a much needed advancement from previous literature relating to the presentation of static images within experimental (i.e., tightly controlled, non-naturalistic) conditions (e.g., Holmes & Zanker, 2012; Nummenmaa et al., 2006) by providing a dynamic visual environment for our participants to explore.

We retain elements of experimental control by employing a counterbalanced design, with five versions of the web page shown to ensure that each post appears in each position upon the web page, allowing us to separate the effect of content from position on the page to some extent. We average our oculomotor metrics (time to first fixation, total fixation duration, and number of fixations) so that for each post we average the oculomotor metrics across the five positions. Given the novel visual environment this provides, it is worthwhile to characterise the distribution of attention across the stimulus whilst participants browse our scroll-able content (Djamasbi, 2014). As fixations provide a correlate of attention (Hayhoe & Ballard, 2005), we map the distribution of total fixation duration across all participants upon an example stimulus. This is represented in the form of a gaze opacity map, which we then interpret to report upon how participants generally explore our stimuli.

We predict that our overall regression model (time to first fixation, fixation duration, fixation number) will be a significant predictor of the emotional rating given to a social media post. We predict a significant positive correlation between the emotional rating of the post and the number of fixations upon the post. We also predict that there will be a positive correlation between the emotional rating of a post and the amount of time spent fixating upon (total fixation duration) the post. Finally, we predict that there will be a negative correlation between time-to-first fixation and emotional rating, with participants, on average over trials, being faster to fixate upon the images that they prefer (e.g., due to quickly progressing through less interesting/preferred items).

## 3.3 Methods

### 3.3.1 Participants

Due to delays induced by addressing the issues highlighted in piloting, data collection was conducted for the five weeks before the university term ended, after which collection was stopped. This resulted in 38 undergraduate university students ranging from 18-31 years old ($M_{age} = 20.35$, $SD_{age} = 3.05$, 33 female), who took part in exchange for course credits or a monetary reward (5 GBP). They all had normal or corrected to normal vision. Three participants (all female) were excluded from the analysis due to having less than 80% valid samples within the eye tracking data. This resulted in the sample size of 35 (each of whom rate five stimuli, providing 175 observations), which assuming an average *apriori* effect size ($f^2 = 0.4$ / $r = 0.2$) gives us a statistical power above the 0.8 recommended (Brysbaert, 2019). Ethical approval was granted from the Royal Holloway Research Ethics Committee.

### 3.3.2 Eye Movement Data Collection

All stimuli were presented on a 23-inch TFT monitor at 1920 x 1080-pixel resolution. The viewing distance was 60cm, and stimuli were presented full-screen using the Internet Explorer browser. Recording took place in a windowless room, ambient lit with fluorescent lighting. Eye movement data was collected using the Tobii TX300 eye tracker at a rate of 300 samples per second. Using Tobii Studio software to present the stimuli and collect eye movement data, a five-point calibration procedure was followed, with manual visual inspection of calibration accuracy for each participant. The raw oculomotor data was analysed using Tobii Studio software, with events detected using an IVF filter (Window length: 20ms, Velocity threshold: 30°/s) and classified as fixations or saccades. Participants were included if the calibration accuracy passed visual inspection, and they had greater than 80% valid eye tracking samples (three exclusions).

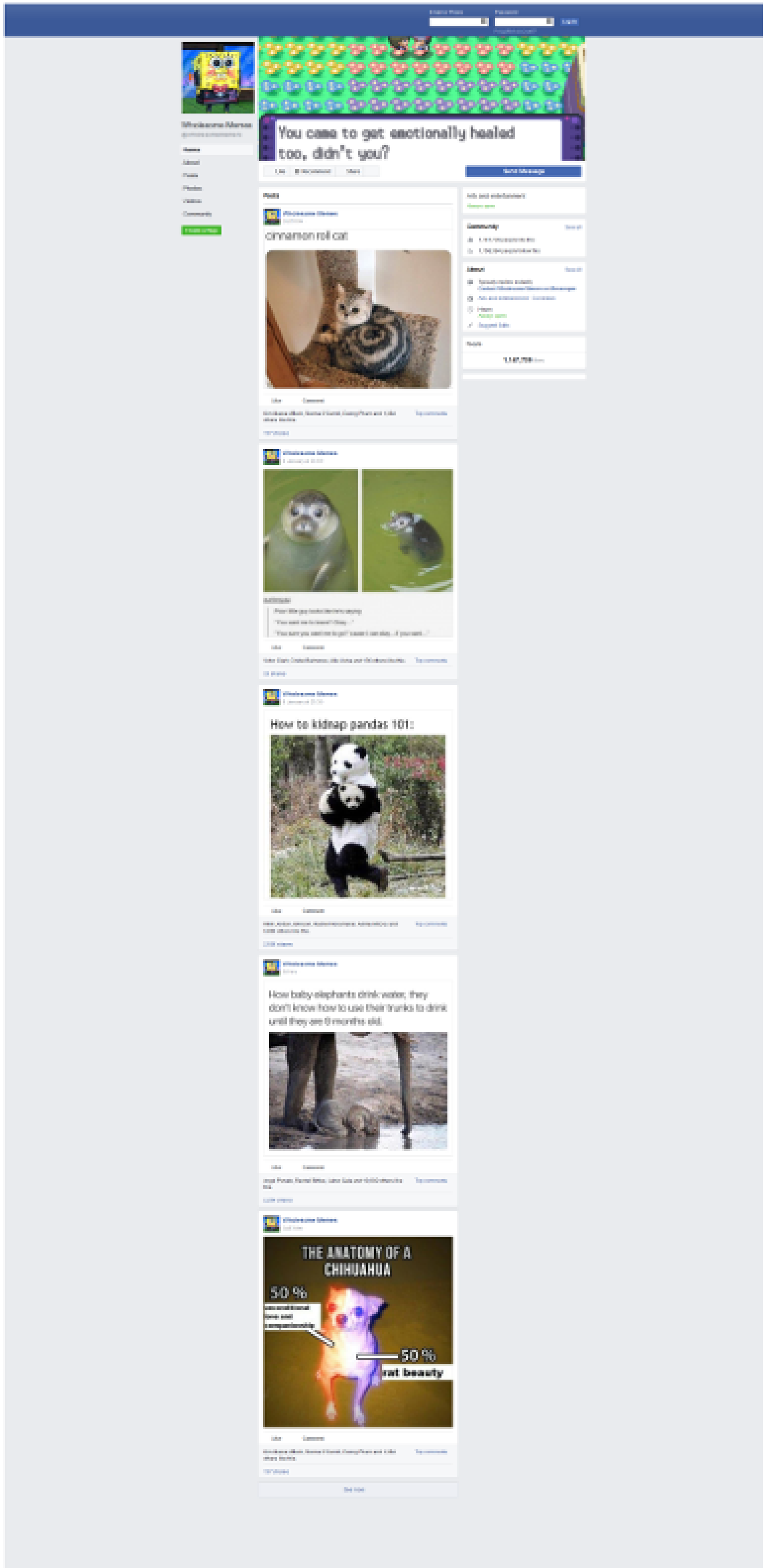


FIGURE 3.1: Stimuli: Animal composite web page

### 3.3.3 Materials

Several methodological challenges had to be overcome to collect and directly compare the visual behaviour of participants while they viewed a social media themed web page. Firstly, social media web pages rely heavily upon a scrolling mechanism. This mechanism creates a visual search paradigm, whereby the participant must choose between viewing the currently displayed selection of visual content, or scrolling in search of new and potentially more rewarding content. As such, to be ecologically valid, our stimuli must replicate this environment. This required that we design a partially occluded stimulus, where not all images were visible on the screen at once, which allowed for the participant to explore via scrolling.

An early approach was to download an example Facebook web page in its entirety and show this within a stimulus presentation software ("Tobii Studio"). However, we encountered technological barriers during the piloting phase, namely, that there was a large offset between the initial loading of the web page and the subsequent display of the visual scene. This delay varied with the central processing unit (CPU) load, leading to nonsystematic variance across (piloting) participants. Furthermore, the action of scrolling caused a similar lag between user input and the visual scene updating. To reduce the computational overhead, we simplified the web page to host a large image within a HyperText Markup Language (HTML) table. This substantially reduced load time and increased responsiveness, mostly by reducing the CPU load associated with dynamic elements. The compromise was relatively small in terms of ecological validity, for example, the side bars no longer track along the page as the participant scrolled. During piloting, we discovered participants did not consciously notice this difference.

Using the above method, we developed the stimuli used within this analysis (Figure 3.1). The web page contains five centrally located composite pictures ('posts'), each containing animals and text. Each post is 13.2cm (12.55° degrees visual angle) horizontally and between 15-18cm (14.25° - 17.06°) vertically, and was assigned as a region of interest (ROI). An example of an annotated web page is given in Figure 3.2. We create five versions of this web page, each time varying the position of the five images (rotating the order on the page) to counterbalance their positions such that in each version every post is located in a unique location in the page (see appendix for stimuli). The images used were selected to be in keeping with the popularity of 'meme' style content upon social media sites, with each displaying an animal and some overlaid text. We allowed our images to vary in their low-level saliency (e.g., contrast, brightness) as this mimics how images would occur in the ecological situations of browsing within a social media website, and we wished our findings to be applicable within this context.

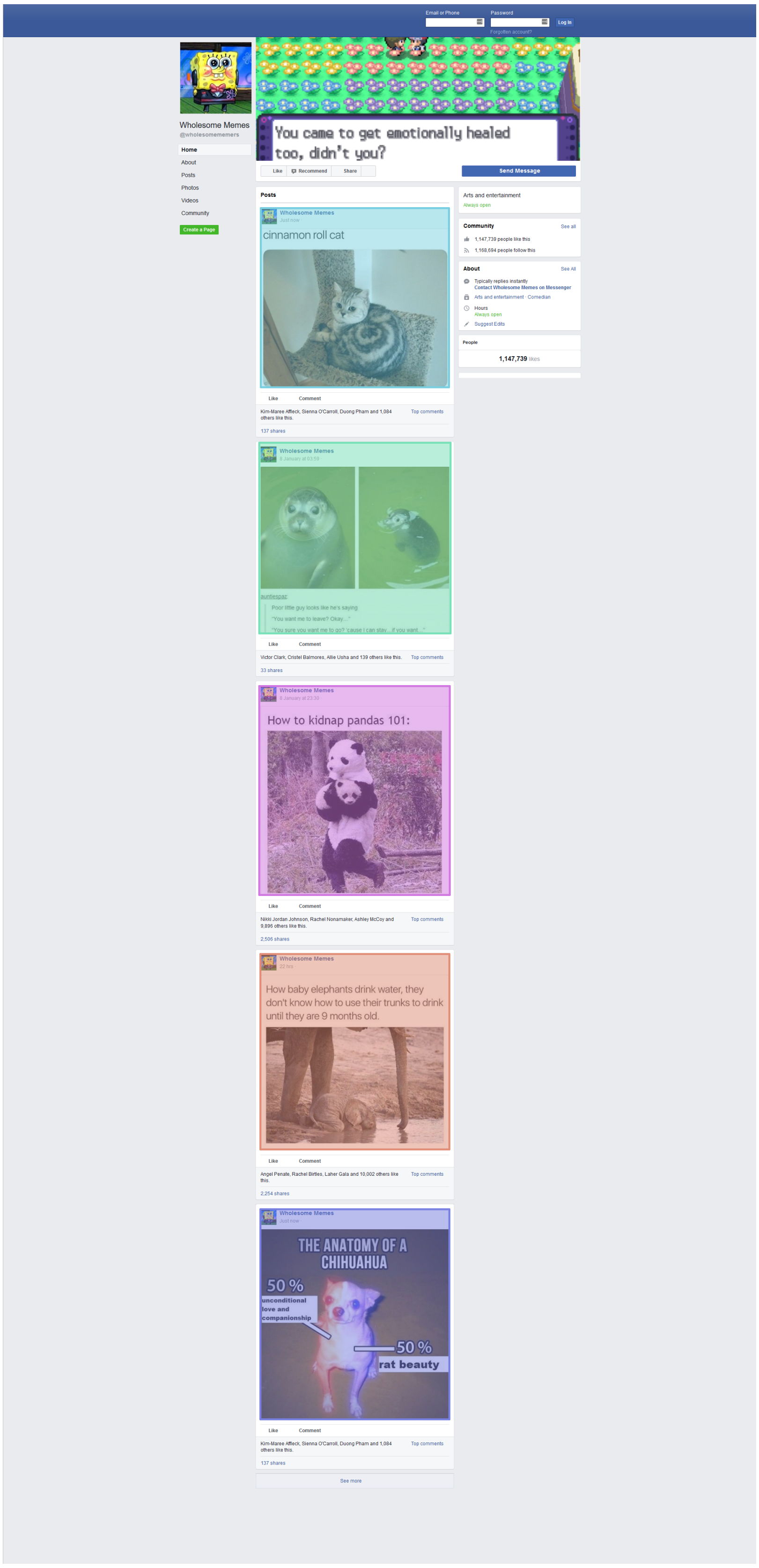


FIGURE 3.2: Stimuli: Animal composite web page with regions of interest highlighted

### 3.3.4 Procedure

Participants first were briefed and given time to provide informed consent before beginning the eye tracking segment. After calibration, participants viewed and scrolled through a demonstration page to ensure familiarity when interacting with the stimulus. After a verbal confirmation of familiarity with the paradigm, written instructions appear upon the page, instructing the participant to 'view the following pages as you would normally'. This was followed by a 30s presentation of an adapted university Facebook page as part of another study, not reported here (see: 4). The stimuli analysed in this paper are then displayed (e.g., the five versions of Figure 3.1, provided in supplementary materials A). We present each of the five versions of this stimulus for 30s, such that every participant sees each post at each position in the page. The display order is randomised across participants, and each web page stimulus is interleaved with fixation cross images located centrally (5s presentation), which the participant was instructed to look at. Specifically, participants were not told that they would be viewing the same content five times to minimises the certainty of what might come next. This was because participants knowing what content is likely to come next will affect the concept of "exploring" the web page, which is undesirable.

After the eye tracking segment was complete, participants rated their subjective emotional valence toward each of the five posts. This consisted of five analogue slider scales (Range: -1 'negative' to +1 'positive', centred on 0) for which participants reported their positive or negative affiliation ('How did viewing this image make you feel?') with each post. Further demographic, personality, and trait characteristics were also collected for use in a machine learning investigation not reported here. This included the NPI-16 narcissism questionnaire (Ames, Rose, & Anderson, 2006), the Rosenberg 10-item self-esteem questionnaire (Robins, Hendin, & Trzesniewski, 2001), and the 60 item NEO Five-Factor Inventory (Costa, 1996).

### 3.3.5 Analysis

For each participant, the mean total fixation duration and number of fixations will be calculated from across the five viewings for each of the five central posts using a region-of-interest approach (see last paragraph of materials for descriptions of ROI). Thus, each of these metrics is the average over all possible positions for the given post. The time-to-first-fixation metric will be calculated in the same way, with missing values (where the participant never views the stimulus) imputed with the trial length of thirty seconds. This results in three metrics describing each participants average viewing behaviour in correspondence with each of the five ROI (35 participants, 5 ROI for a total of 175 observations). These metrics will be entered into a linear regression model predicting the affective rating of the contents of the ROI. Thus, whilst participants are likely to catch on to the content being the same as trials progress (which changes the context of their viewing behavior) the counter-balancing design means that the effect of this is minimised by averaging across the participants. Scrolling behaviour was recorded but not analysed in this study.

## 3.4 Results

### 3.4.1 Visualising Viewing Behaviour

To gain a general insight into the viewing patterns of participants upon this novel stimulus, we calculate a gaze opacity map. Here, lighter ("bright") regions indicate a greater proportion of the participant's viewing duration was spent fixating upon that region, while darker regions indicate locations where few, if any, participants spent time fixating. This effectively highlights which parts of the visual content were commonly attended to, along with regions that did not commonly receive overt visual attention.

We collect the visual behaviour of all participants (N=35) upon one version of the web page (randomly chosen). Note that this may be the first presentation of this web page for some participants, but the fifth for others due to the presentation order being randomized across participants. We represent the distribution of total fixation duration across all participants for the 30 seconds of viewing duration by applying a smoothing process (Gaussian kernel of 50 pixels) upon the aggregated total fixation duration values. We then scale the gaze opacity image such that the brightest (e.g., the most opaque) region represents the peak fixation duration (19.83 seconds), and the darkest black (e.g., the least opaque) the minimum (Figure 3.3). As a short methodological note, more complex visual stimuli often require longer to process (Becker, 2011). As such, the representation of a region in a gaze opacity map is linked to both the density of information present within that visual region, alongside its visual saliency. Thus, the results of such visualisations should not be interpreted as a direct correlate of visual salience.

From visual inspection of the gaze opacity map, we can see that participants are very efficient at picking out information in the page, as very little time is spent viewing elements of the web page that are invariant across the social media site or lack content. Examples include the blue ribbon at the head of the web page and the grey regions to the side of the central post elements. Our visual attention is biased towards the center of the visual scene, often referred to as the central tendency bias (Tatler, 2007). This may have affected our results as the key content in the web page is presented within a single, centrally located, column. By having this layout the web page design takes advantage of the central tendency bias and thus our results may over-estimate how much time would usually be spent on each image were they not centrally located. There also appears to be a strong bias for the face region of each piece of content (even if this is an animal face).

In summary, we may characterise our participant's viewing behaviour as a central biased search of social media content that is efficient in picking out each post's content, and identify that participants clearly reach, and are able to spend time fixating, the bottom image within the presentation duration.

### 3.4.2 Descriptive Statistics

We find that there is little to no mean difference across our stimuli upon the metric of time to first fixation, with an equal amount of variance across all stimuli. Across viewings, our participants on average spent longer looking and displayed the most fixations at the post (e.g., region of interest) displaying a seal than any other post. In

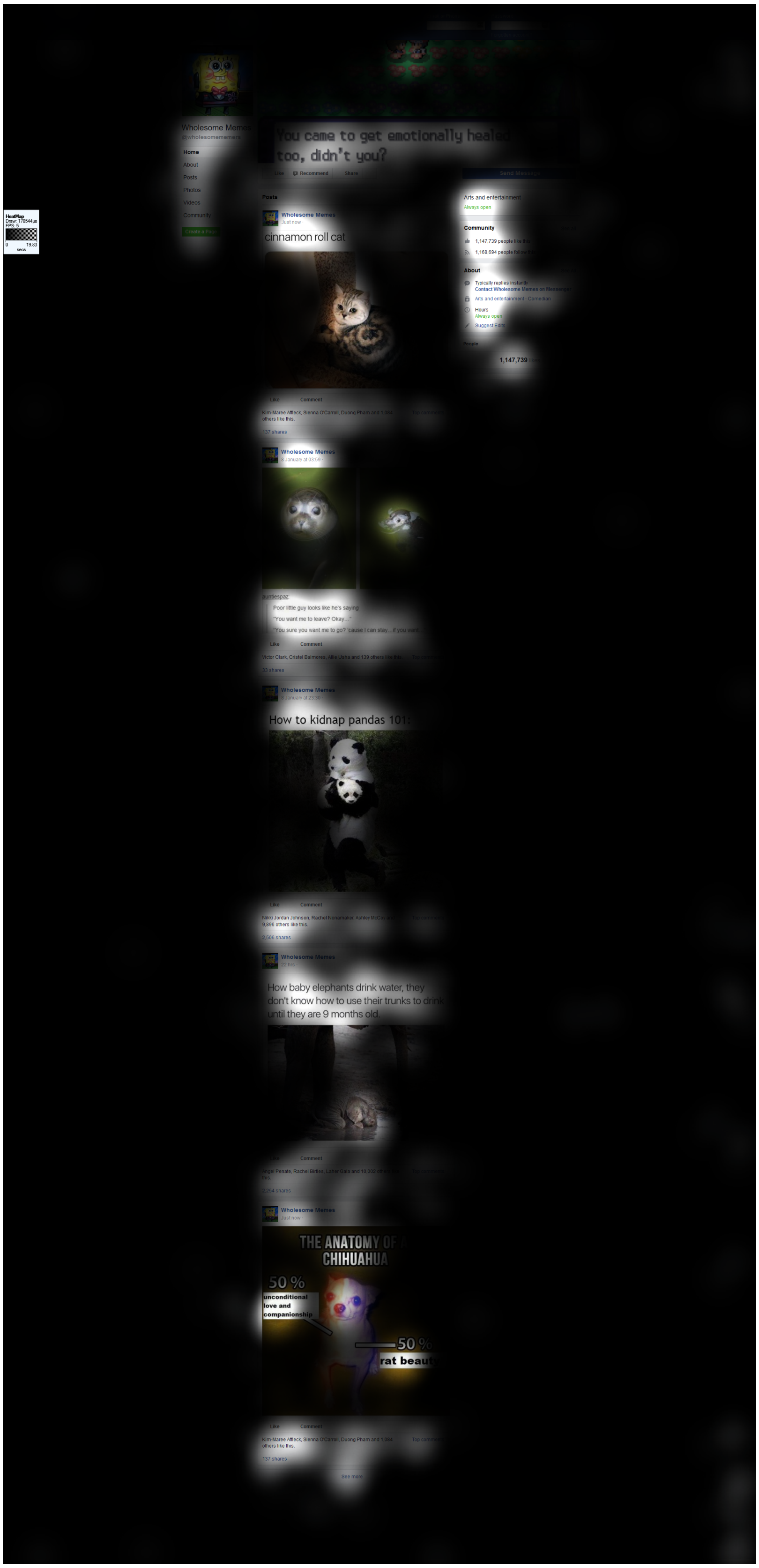


FIGURE 3.3: Gaze Opacity Map: Total fixation duration upon animal based Facebook page. 35 participants, Gaussian smoothing kernel (50px), Opacity scaled to maximum absolute fixation duration (19.83s).

contrast, the stimulus displaying a cat received the least fixations and fixation duration than any other stimuli (see: Table 3.1). This indicates that the different stimuli in each post have evoked varying oculomotor behaviour, and we now investigate whether the individual variation across participants in this behaviour is informative of emotional valence score.

TABLE 3.1: Descriptive Statistics For Visual Metrics: Grand Mean and Standard Deviation (SD) for metrics describing visual behaviour toward central posts upon the social media themed web page. Time to first fixation and total fixation duration are reported in seconds.

| | **Time to First Fixation** | **Total Fixation Duration** | **Number of Fixations** |
|---|---|---|---|
| | **Mean (SD)** | **Mean (SD)** | **Mean (SD)** |
| **Cat** | 12.23 (3.17) | 11.23 (3.17) | 43.54 (8.96) |
| **Dog** | 12.15 (3.24) | 14.94 (4.26) | 59.69 (11.94) |
| **Elephant** | 12.2 (3.57) | 15.38 (3.77) | 67.49 (16.01) |
| **Seal** | 11.87 (3.46) | 17.08 (4.54) | 71.91 (17.8) |
| **Panda** | 12.1 (3.82) | 11.99 (3.88) | 49.91 (13.2) |

### 3.4.3 Main Results: Does visual behaviour predict emotional valence rating?

We first check whether our data meets the assumptions for a multiple linear regression analysis. The distribution of our outcome variable of emotional rating failed the Shapiro-Wilks test of normality (p<.001). However, the distribution of the residuals is not significantly different from normal, and a Goldfeld-Quandt test shows that our model still meets the requirement of homoscedasticity ($GQ(84, 83) = 1.22, p = .181$). Furthermore, there is no evidence of multicollinearity (Value Inflation Factor < 3.62, Zero Order Correlations in Table 3.2) or auto correlations in the data (Durbin-Watson's $W = 1.79, p = 0.182$). As such, it is appropriate to continue with the analysis. Our overall model (Equation 3.1) explained 4.8% of the variance in emotional valence ratings (adjusted $R^2 = 0.032$), and was a significant predictor ($F(3, 171) = 2.905, p = .036$). Time to first fixation (TTFF) upon the stimuli (Figure 3.4) was a significant predictor of emotional rating ($\beta = -.028, p = .015$), with a

.028 reduction in predicted rating per second taken to first fixate the stimulus. Total fixation duration (TFD) and number of fixations (NF) upon the stimulus were not significant predictors of emotional rating ($p > .168$).

TABLE 3.2: Zero Order Correlations

| | 1 | 2 | 3 |
|---|---|---|---|
| **1. Time to First Fixation** | - | -.172 | -.135 |
| **2. Total Fixation Duration** | - | - | .851 |
| **3. Number of Fixations** | - | - | - |

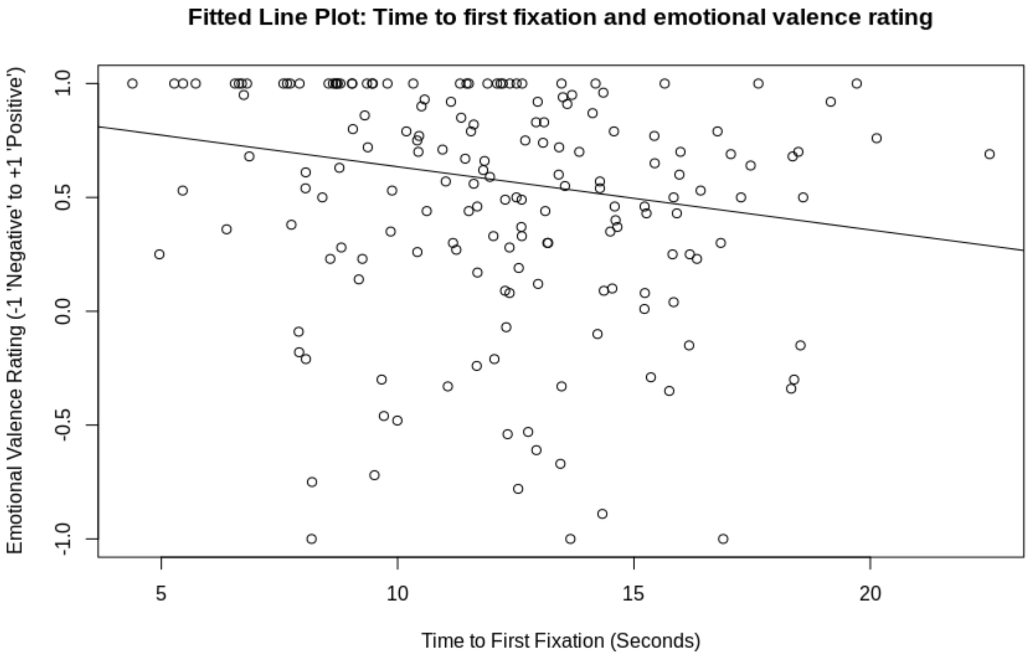


FIGURE 3.4: Line of best fit for time to first fixation and emotional valence rating

### 3.4.4 Exploratory Analysis: Visual behaviour and arousal

As it has been suggested that it may be the magnitude (rather than the direction) of the emotional reaction that leads to preferential viewing behaviour (Nummenmaa et al., 2006), we repeat the above analysis upon the absolute values of the outcome variable. We refer to this as the 'absolute model'. To illustrate, a participant who rated one post as mildly unpleasant (e.g., -0.2), and another as mildly pleasant (e.g., 0.2) would now have the same score (e.g., 0.2) for both ratings. If we find that the results are improved, this supports the claim that it is the magnitude rather than the direction of the emotional reaction that draws visual attention.

We find that after transformation (taking the absolute value), our outcome variable is still not normally distributed ($W = 0.912, p < .001$). The remaining assumptions were met, with homoscedasticity ($GQ(84, 83) = 0.68, p = .961$), no significant multicollinearity ($VIF < 3.7$) or auto-correlations ($DW = 1.92, p = .552$). Overall, the absolute model (shown in Equation 3.2) explained 10% of the variance in absolute emotional valence ratings (Adjusted $R^2 = 0.085$) and was a significant predictor ($F(3, 171) = 6.36, p < .001$). The variables of time to first fixation ($\beta = -.022, p < .001$), total fixation duration ($\beta = -.030, p = .002$) and number of fixations ($\beta = .006, p = .010$) were all significant predictors of the absolute emotional valence score. We graph the result for time to first fixation to allow comparison of signed values in Fig 3.4 to the unsigned values in Figure 3.5.

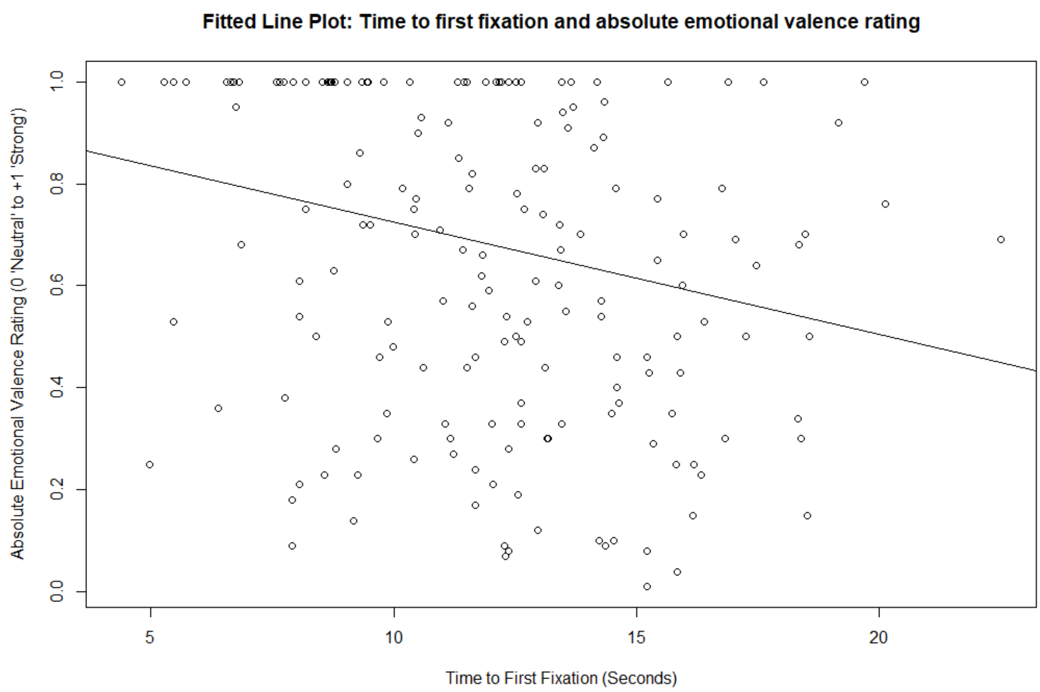


FIGURE 3.5: Line of best fit for time to first fixation and absolute emotional valence rating

$$Rating = 0.912 + -.028TTFF + -.031TFD + .001NF \quad (3.1)$$

$$|Rating| = 0.944 + -.022TTFF + -.030TFD + .006NF \quad (3.2)$$

FIGURE 3.6: Multiple Linear Regression Equations: Emotional Rating (3.1) and Absolute Emotional Rating (3.2).
TTFF: Time to First Fixation. TFD: Total Fixation Duration. NF: Fixation Number

## 3.5 Discussion

Studies investigating gaze allocation using static pictures have previously failed to replicate when investigated in ecologically valid, non-laboratory settings (Tatler et al., 2011). As such, in this paper we investigated the relationship between visual behaviour and affective judgement responses while participants browsed a novel stimulus; a mock social media web page. Through visual inspection of the gaze opacity map, we found that our social media themed web page elicited visual behaviour that can be well characterised as being focused upon the exploration of the central elements ('posts') which were the subject of our experiment, and that participants were able to fully explore each of the five posts within the given viewing duration. The efficiency of the visual search is likely aided by a familiar social media layout that predictably displays content within the central region. However, future research will be needed to investigate whether the influence of additional distractions such as advertisements and user-interactive elements (not present in this experiment) significantly impacts this general pattern of viewing behaviour. The finding that oculomotor behaviour is biased towards, and influenced by, the presence of facial regions in visual stimuli is in keeping with previous literature (Armstrong & Olatunji, 2012; Macdonald & Tatler, 2018), and we note that each of our five central images is homogeneous in being animal themed and showing the face, which prevents this from being a substantial confound.

To a certain extent, we find that our main regression results (predicting emotional rating) are in keeping with the previous literature investigating the association between visual behaviour and affective responses upon static images (Maughan et al., 2007; Nummenmaa et al., 2006; Shimojo, Simion, Shimojo, & Scheier, 2003). We find that together the three oculomotor metrics (time to first fixation, total fixation duration, and number of fixations) explain a statistically significant amount of variance in the participant's affective judgement responses (ranging from -1 'negative' to +1 positive). However, as the emotional rating model only accounts for a small amount of variance (4.8%) the model is unlikely to provide accurate predictions that are useful within applied settings (e.g., such as market research). In keeping with our hypothesis that participants will choose to preferentially view images that they prefer, first, we find the metric of time-to-first-fixation to be a significant negative predictor of emotional rating. This expands upon the findings of earlier research which found preferred stimuli to be viewed earlier in static image paradigms (Nummenmaa et al., 2006), by demonstrating that this association holds even when the stimulus is not initially visible. As such, our results suggest that it is not just that participants preferentially fixate the objects they prefer (e.g., have positive emotional associations with) within the immediately available visual scene, they additionally display preferential scrolling behaviour (e.g., rapidly moving on from less preferred stimuli) resulting in faster time to first fixation upon stimuli even when the stimulus is occluded. However, as our repeated design led to the participant being able to anticipate the content of the web page in later trials, we note that this finding may only be applicable within situations where the participant has a strong prior expectation of the web site's content.

The directional hypotheses for the metrics of total fixation duration and number of fixations were unsupported within our main results, with non-significant effects being found for both metrics. This contrasts with previous research upon static images, which has found the preferred content to be fixated more frequently and for longer (Holmes & Zanker, 2012; Maughan et al., 2007). Given this, one may initially conceptualise our main results as suggesting that the positive association between fixation duration and number and positive emotional evaluation may not always occur while participants browse content within the context of a social media themed web page environment. As such, looking (e.g., exhibiting more fixations or spending a longer duration upon a stimulus) does not always equate to a subject liking (e.g., giving a stronger positive appraisal to) a particular stimulus while browsing online.

Interestingly, in our exploratory analysis we found that the same prediction variables, with an almost identical set of coefficients, could explain twice as much variance (10%) when predicting absolute emotional valence scores (e.g., ranging from 0-1) rather than emotional valence scores. This supports the suggestion by Nummenmaa et al. (2006) that it is the magnitude of the emotional reaction rather than the direction (e.g., positive or negative) that influences visual attention. A caveat to this is that the emotional valence measures were biased toward positive values (with very few observations in the negative values). This may have led to the model being biased towards predicting positive values, which was alleviated by taking the absolute values. This would suggest that either our result of absolute emotional valence being the better predictor may be an artifact of our distribution of emotional valence scores, or that we may indeed be underestimating the magnitude of the difference between predicting absolute versus signed values. To elaborate, the model predicting the signed values may be performing better than expected due to the sparsity of negative values (which would otherwise reduce the fit of the signed model, and increase the positive benefit of predicting the absolute values). As our finding is in keeping with previous literature that has investigated a greater range of negative images and found that unpleasant pictures also capture visual attention (Nummenmaa et al., 2006; Ono & Taniguchi, 2017), we believe that the latter is the most likely scenario. Even so, a direction for future research is to replicate this experiment with a different stimulus set with a more distributed range of valences to ensure the results are not related purely to the distribution of the valence scores.

For the absolute model, all three oculomotor metrics were significant predictors, but the total fixation duration was not in the expected direction, as it was negatively associated with absolute emotional rating. As such, by bringing together results from the main and exploratory analysis, it appears that visual behaviour within a social media themed web page presents a unique pattern of association with emotional ratings (i.e., preference decisions) compared to results from static image paradigms. Namely, the finding that preferred, or simply more emotionally arousing (e.g., the absolute magnitude regardless of being positive or negative), content is also fixated for longer (e.g., as in; Holmes & Zanker, 2012; Maughan et al., 2007; Nummenmaa et al., 2006; Shimojo et al., 2003) does not seem to generalise to our web page style visual stimulus. We reach this conclusion because total fixation duration was not a significant predictor when looking at emotional valence and was significantly negatively associated with the absolute emotional valence score; both of these findings being disparate from that reported by the above literature.

A candidate for this disparity in findings is the different methods of stimulus presentation employed. In previous literature, static images were either presented individually (Maughan et al., 2007), simultaneously on either side of a fixation cross (Shimojo et al., 2003) or simultaneously presented radially around a fixation cross (Holmes & Zanker, 2012). Simultaneous presentation allows an equal opportunity for the participant to first fixate upon any of the available images and distribute their gaze as they so desire. However, in our web page design, not all stimuli are simultaneously available, as the scrolling mechanism leads to the full or partial occlusion of competing stimuli. This accurately reflects real-world social media design, and has an impact upon visual behaviour as the participant now finds themselves with a novel decision; whether to exploit the current scene or explore the currently occluded stimuli by scrolling. As such, participants, rather than the experimenter, decide upon the viewing duration for each segment and may choose to move on rather than continue fixating upon previously viewed content. Within a traditional static image design, participants would be unable to make this decision as they are constrained by the preset experimental viewing duration (e.g., they are forced to continue viewing until a new stimulus is shown). This may lead participants to continue to view the preferred stimulus. As such, we suggest that the difference in findings regarding total fixation duration may be tied to the manner of interacting with the content upon the screen. If we were to alter our paradigm by having the screen continually scrolling outside of the participant's control, this may replicate the association found between total fixation duration and preference as found in previous literature. This was not conducted as it breaks from the ecological validity of the current design, but is a direction for a future research.

This has important implications for market research studies using eye tracking techniques to evaluate the emotional impact of social media advertisements, as our results suggest that showing the advert within a naturalistic social media web page context may lead to researchers underestimating the participants emotional reaction to the content if they use total fixation duration as an index of emotional salience. This is because, in this context, our results suggest that researchers can not assume that the participant will look longer at items that evoke an emotional reaction. In the same manner, the researcher may also overestimate the amount of time that the individual will spend fixating upon the advertisement when it is presented within realistic social media settings if they extrapolate from measurements of total fixation duration gained within a static image based paradigm.

### 3.5.1 Limitations

The finding in our exploratory analysis that absolute emotional valence is positively predicted by the number of fixations, yet negatively predicted by total fixation duration, is worthy of consideration. The most straightforward explanation, as we presented composite images of an animal themed nature overlaid with text, is that participants who had a stronger emotional reaction to the image spent their time reading the associated text (or, vice versa, that reading the associated text increased the absolute subjective rating of the stimulus). Reading evokes a visual pattern characterised by frequent short fixations intersected by small saccadic movements to the next section of letters, which would produce a greater number of shorter fixations (Carrasco, 2011). Thus, when the fixation duration is summed across all fixations, this will not be reflected in the total fixation duration metric. However, the finding that total fixation duration is negatively associated within the exploratory analysis is a finding that requires further exploration. This is a potential limitation of our study, as we did not control the amount of text (or the image characteristics) and it may be that the association between total fixation duration and absolute emotional response varies with content type. While our stimulus is representative of social media content, in that the content is not controlled for physical salience properties, a possible direction for future research would be to replicate our experiment in social media themed web pages that contain content of a single category (e.g., contain purely image, text or video based media) to address this ambiguity.

Finally, we note that while statistically significant, the overall predictive power of our models is relatively weak and unlikely to yield accurate predictions of emotional valence or arousal. Future research may wish to employ more sophisticated pattern recognition techniques, to evaluate whether non-linear statistical associations can provide additional insight when predicting emotional valence from visual behaviour upon social media style content.

### 3.5.2 Conclusion

In summary, our research supports that visual behaviour reflects preference decisions in social media themed web pages. In particular, we find that the visual behaviour metrics investigated best reflect absolute emotional valence rather than whether the stimulus was perceived positively or negatively. We also find the relationship between total fixation duration and (absolute) emotional valence to be unique from that suggested by the literature based upon static images. This has important implications for the use of eye tracking in market research upon social media style stimuli, where results from statically presented images may not generalise well to how participants will engage with the visual stimulus when presented *in-situ*. Finally, we acknowledge that additional non-linear associations between eye movement metrics (e.g., time to first fixation) and individual differences in preference may exist that were not captured within this study. To address this shortcoming, future research may wish to pursue the relationship between a greater range of metrics describing visual behaviour and preference decisions using pattern recognition techniques within a machine learning approach.

## References

Ames, D. R., Rose, P., & Anderson, C. P. (2006). The NPI-16 as a short measure of narcissism. *J. Res. Pers. 40*(4), 440–450. doi:10.1016/j.jrp.2005.03.002

Anderson, B. (2013). A value-driven mechanism of attentional selection. *J. Vis. 13*(3), 7–7. doi:10.1167/13.3.7

Armstrong, T., & Olatunji, B. O. (2012). Eye tracking of attention in the affective disorders: A meta-analytic review and synthesis. *Clin. Psychol. Rev. 32*(8), 704–723. doi:10.1016/j.cpr.2012.09.004

Becker, S. I. (2011). Determinants of dwell time in visual search: Similarity or perceptual difficulty? *PLoS One*, *6*(3), e17740. doi:10.1371/journal.pone.0017740

Brysbaert, M. (2019). How Many Participants Do We Have to Include in Properly Powered Experiments? A Tutorial of Power Analysis with Reference Tables. *Journal of Cognition*, *2*(1), 16. doi:10.5334/joc.72. pmid: 31517234

Carrasco, M. (2011). Visual attention: The past 25 years. *Vision Research*, *51*(13), 1484–1525. doi:10.1016/j.visres.2011.04.012

Chelazzi, L., Perlato, A., Santandrea, E., & Della Libera, C. (2013). Rewards teach visual selective attention. *Vision Res. 85*, 58–62. doi:10.1016/j.visres.2012.12.005

Costa, P. T. (1996). Work and personality: Use of the NEO-PI-R in industrial/organisational psychology. *Appl. Psychol. 45*(3), 225–241. doi:10.1111/j.1464-0597.1996.tb00766.x

della Libera, C., Perlato, A., & Chelazzi, L. (2011). Dissociable effects of reward on attentional learning: From passive associations to active monitoring. *PLoS One*, *6*(4), e19460. doi:10.1371/journal.pone.0019460

Djamasbi, S. (2014). Eye Tracking and Web Experience. *AIS Transactions on Human-Computer Interaction*, *6*(2), 37–54.

Hayhoe, M., & Ballard, D. (2005). Eye movements in natural behavior. *Trends Cogn. Sci. 9*(4), 188–194. doi:10.1016/j.tics.2005.02.009

Holmes, T., & Zanker, J. M. (2012). Using an Oculomotor Signature as an Indicator of Aesthetic Preference. *i-Perception*, *3*(7), 426–439. doi:10.1068/i0448aap

Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nat. Rev. Neurosci. 2*(3), 194–203. doi:10.1038/35058500

Kelley, T. A., & Yantis, S. (2010). Neural Correlates of Learning to Attend. *Front. Hum. Neurosci. 4*. doi:10.3389/fnhum.2010.00216

Kowler, E. (2011). Eye movements: The past 25years. *Vision Res. 51*(13), 1457–1483. doi:10.1016/j.visres.2010.12.014

Macdonald, R. G., & Tatler, B. W. (2018). Gaze in a real-world social interaction: A dual eye-tracking study. *Q. J. Exp. Psychol.* 174702181773922. doi:10.1177/1747021817739221

Maughan, L., Gutnikov, S., & Stevens, R. (2007). Like more, look more. Look more, like more: The evidence from eye-tracking. *J. Brand Manag. 14*(4), 335–342. doi:10.1057/palgrave.bm.2550074

Niu, Y., Todd, R. M., Kyan, M., & Anderson, A. K. (2012). Visual and emotional salience influence eye movements. *ACM Trans. Appl. Percept. 9*(3), 1–18. doi:10.1145/2325722.2325726

Nummenmaa, L., Hyönä, J., & Calvo, M. G. (2006). Eye movement assessment of selective attentional capture by emotional pictures. *Emotion*, *6*(2), 257–268. doi:10.1037/1528-3542.6.2.257

Ono, Y., & Taniguchi, Y. (2017). Attentional Capture by Emotional Stimuli: Manipulation of Emotional Valence by the Sample Pre-rating Method. *Jpn. Psychol. Res. 59*(1), 26–34. doi:10.1111/jpr.12142

Robins, R. W., Hendin, H. M., & Trzesniewski, K. H. (2001). Measuring global self-esteem: Construct validation of a single-item measure and the Rosenberg Self-Esteem Scale. *Personal. Soc. Psychol. Bull. 27*(2), 151–161. doi:10.1177/0146167201272002

Shimojo, S., Simion, C., Shimojo, E., & Scheier, C. (2003). Gaze bias both reflects and influences preference. *Nat. Neurosci. 6*(12), 1317–1322. doi:10.1038/nn1150

Tatler, B. (2007). The central fixation bias in scene viewing: Selecting an optimal viewing position independently of motor biases and image feature distributions. *J. Vis. 7*(14), 4. doi:10.1167/7.14.4

Tatler, B., Hayhoe, M., Land, M., & Ballard, D. (2011). Eye guidance in natural vision: Reinterpreting salience. *J. Vis. 11*(5), 5–5. doi:10.1167/11.5.5

Tatler, B. W., Wade, N. J., Kwan, H., Findlay, J. M., & Velichkovsky, B. M. (2010). Yarbus, eye movements, and vision. *Iperception. 1*(1), 7–27. doi:10.1068/i0382

# Chapter 4

# Looking into web browsing: Using eye movements whilst browsing social media type content to predict personality traits

## 4.1 Abstract

Digital records of our behavior on social networking sites (SNS) can provide insight into our personal traits and values. An aspect of our behavior that has the future potential to be monitored on a large scale are the eye movements we make. Eye movements partly reflect the influence of our cognitive biases through endogenous attention, and visual behavior is influenced by social and emotional cues, as well as visual properties of an image. Machine learning techniques are a powerful method for detecting patterns using multiple eye movement parameters as input. As such, we explore whether eye movements while browsing SNS are informative of personality. We tracked the eye movements of thirty-five participants as they engaged with two different social media inspired web pages. Using only visual behavior, our models predicted the personality traits of Extroversion (71%), Agreeableness (71%), and Neuroticism (69%) significantly above chance. We discuss the relative contribution of describing overall physiological parameters (based on overall fixation/saccade distributions), or attentional allocation (location specific descriptors), upon classifier performance. Our findings highlight that eye movements may reveal more about

personal information than is currently of public knowledge.

## 4.2 Introduction

Our activity upon social networking sites (SNS) provides a digital account of how we express ourselves to others, convey our preferences, and respond to social situations – with each of these being influenced by our personal characteristics (Nadkarni & Hofmann, 2012). This provides a rich source of online data that is well suited to exploration with pattern recognition techniques, and there has been a focus upon how personal beliefs, values, and traits are reflected in, and can be predicted from, recordings of human behavior when using social media (Bleidorn & Hopwood, 2018). The recent surge of interest in the application of machine learning principles to psychological research (for a review, see: Adjerid & Kelley, 2018) has identified that a broad range of outcomes, such as an individual's psychological wellbeing and intelligence (Settanni, Azucar, & Marengo, 2018), can be predicted from human behavior upon SNS. An early example is the work of Kosinski, Stillwell, and Graepel (2013), who illustrated that a user's past history of 'liking' pages upon the social media website Facebook was sufficient to predict a user's sexuality (88% accuracy) and political preference (85%) significantly better than chance. Furthermore, the accuracy of predictions made upon certain aspects of their personality (Openness to Experience) showed comparable reliability to that of established psychological questionnaires (Kosinski et al., 2013).

The interest in understanding how aspects of an individual's personal traits and attributes are reflected in their digital footprint (Weaver & Gahegan, 2007) comes from academic and commercial sources, with potential applications in advertisement (Matz, Kosinski, Nave, & Stillwell, 2017; Matzler, Bidmon, & Grabner-Kräuter, 2006), optimising human-computer interaction (Kruijff, Swan Ii, Feiner, Swan, & Feiner, 2010; Stanney, Mourant, & Kennedy, 1998) and as a tool for psychological research into personality (Bleidorn, Hopwood, & Wright, 2017; Kosinski, Matz, Gosling, Popov, & Stillwell, 2015).

Importantly, the ability to accurately predict sensitive information about an individual's personal attributes without users' knowledge should also raise concern over privacy issues. An opinion survey of search engine use found that 68% of users expressed negative reactions to targeted advertising (Purcell, Brenner, & Rainie, 2012); thus, it appears that targeted advertising is often not positively received by the general public. In order for the general public to make informed decisions about sharing their personal data, there is a need for further research (and education) into the range of inferences that can be made from this data. Furthermore, the emergence of computers with inbuilt eye tracking capability combined with progressive improvements in the quality of webcam eye tracking (e.g. Papoutsaki, Laskey, & Huang, 2017), a new form of data that users may be asked to share is their eye movements. In this paper, we aim to address this possibility by exploring whether personality under the big five personality traits (McCrae & Costa, 2004) can be decoded from an individual's visual behavior upon a SNS themed web page.

### 4.2.1 Visual behavior is linked to, and influenced by, emotional responses and cognitive states

It is well known that visual behavior in response to the same stimulus will vary depending upon the task given to a participant, especially within natural scenes (Tatler, Wade, Kwan, Findlay, & Velichkovsky, 2010). As such, the cognitive state of the viewer influences the emerging expression of the oculomotor behavior (Saalmann, Pigarev, & Vidyasagar, 2007). This is often referred to as top-down attention, and is driven by the cognitive biases (Eckstein, Guerra-Carrillo, Miller Singley, & Bunge, 2017) and preferences (Shimojo, Simion, Shimojo, & Scheier, 2003) of the individual participant. The reflection of these cognitive biases can be seen within experiments investigating the orientation of visual behavior toward objects with emotional content. It has been shown that when presented with an emotional image next to a neutral image, participants are more likely to first fixate and to spend longer upon emotive images than neutral images, even when instructed to avoid looking at the emotional content (Nummenmaa, Hyönä, & Calvo, 2006). This behavior is not well described by bottom-up attention, which is driven by the physical properties (e.g., contrast and color) of the visual scene (Itti & Koch, 2001). As such, it has long been proposed that our visual behavior reflects our thoughts and feelings, providing a "window into the soul". Due to the influence of an individual's emotional and cognitive biases, eye movements are influenced by personality factors, such as trait anxiety (Fox, Russo, & Dutton, 2002), and can reflect cognitive processes, such as attention (Armstrong & Olatunji, 2012; Chelazzi, Perlato, Santandrea, & Della Libera, 2013) and mental effort (Eckstein et al., 2017).

### 4.2.2 Personality traits can be predicted from visual behavior

To investigate whether an association exists between the expression of visual behavior and personality under the big five personality model (McCrae & Costa, 2004), Rauthmann, Seubert, Sachse, and Furtner (2012) recorded over two hundred participants' eye movements while they viewed three abstract animations. Each animation was designed to be devoid of meaningful social or topical information, with Rauthmann et al. (2012) proposing that visual behavior, as an autonomic process (McDougal & Gamlin, 2015), should be linked to our personality regardless of stimulus content. The authors focused upon three fixation-based metrics describing the number, mean duration, and dwell time upon each animation. Using a linear mixed modelling design, it was found that these metrics were significantly correlated with trait Openness to Experience, Extroversion and Neuroticism. Rauthmann et al. (2012) interprets this as supporting the hypothesis that there are trait-gaze links, and that these are expressed due to non-volitional links between personality and visual behavior. However, their study also illustrates that not all aspects of an individual's personality may be reflected in visual behavior toward the designed stimuli (i.e., containing little social or affective properties); specifically, the traits Conscientiousness and Agreeableness demonstrated no significant correlation with fixation number, mean duration or dwell time. Furthermore, due to the unnatural animations shown, their findings do not demonstrate whether trait-gaze correlations occur for visual behavior within more ecologically valid settings.

To evaluate whether everyday eye movements are informative of personality traits, Hoppe, Loetscher, Morey, and Bulling (2018) placed mobile eye tracking glasses upon 42 participants as they walked to the local campus shop, purchased some minor items, and returned to complete the NEO-FFI personality questionnaire (McCrae & Costa, 2004). From this, the authors derived 207 measures; for each participant, these described various properties of physiological events (pupil diameter, fixations, saccades, and blinks), the distribution of fixation behavior across the visual scene (eight-by-eight heatmap), and 'n-gram' features describing sequences of physiological events. The authors framed the task as an instance of supervised classification by separating the scores upon each personality trait into low, medium, and high categories using equal frequency binning. Using this data to train random forest classifiers, they were able to classify Neuroticism (40.3%), Extroversion (48.6%), Agreeableness (45.9%) and Conscientiousness (43.1%) above chance (33%). These findings illustrate that eye movements upon stimuli not specifically designed to elicit personality-congruent oculomotor behavior and across a range of tasks, may be useful for predicting aspects of an individual's personality (Hoppe et al., 2018; Rauthmann et al., 2012).

Notably, in contrast to previous literature (Rauthmann et al., 2012), Hoppe et al. (2018) found that the models predicted both traits Agreeableness and Conscientiousness above chance. This could be attributed to key methodological differences, including the task given to participants and the inclusion of over two hundred descriptions of visual behavior within a machine learning paradigm. As certain aspects of personality are partially defined by responses to social and affective situations (Mischel & Shoda, 2008), it is possible that the visual behavior elicited by interacting with other customers and cashier staff within Hoppe et al. (2018) led to the expression of strong statistical trait-gaze relationships that are essential for building classifiers with the ability to accurately predict aspects of an individual's personality from visual behavior. As such, by avoiding stimuli with social and affective properties, Rauthmann et al. (2012) may have underestimated the usefulness of trait-gaze associations. This is especially salient given that specific personality traits, such as Extroversion, are defined by an individual's tendency to seek or avoid social situations (Mischel & Shoda, 2008).

However, the fact that random forest classifiers are non-linear classifiers, which contrasts to the generalized linear model approach of Rauthmann et al. (2012), obscures direct comparisons between the two methods. It is not possible to tease apart whether the additional traits predicted by Hoppe et al. (2018) were enabled via the incorporation of social information in the visual scene, the addition of over 200 additional new descriptions of visual behavior, or the identification of non-linear trends in the data. As such, it remains to be seen if describing visual behavior in response to stimuli with social and affective properties leads to trait-gaze behavior informative of Conscientiousness and Agreeableness. As the most informative features for predicting Conscientiousness and Agreeableness frequently included descriptions of saccadic behavior (not used by Rauthmann), it may be that the random forest approach employed by Hoppe et al. (2018) identified novel trait-gaze associations within these metrics that were not investigated previously. As such, we summarize the work of Hoppe et al. (2018) and Rauthmann et al. (2012) as illustrating that personality is associated with the expression of fixation based visual behavior upon stimuli lacking strong social or affective properties, and that saccadic behavior (in combination with a wide range of oculomotor metrics) supports above-chance performance within classification tasks for a range of personality traits. This illustrates that trait-gaze links are found in a diverse range of stimuli, and that both saccadic and fixation behavior may be useful for predicting personality trait scores.

Recent literature suggests that describing visual behavior in response to stimuli with deliberately selected affective properties makes a dramatic difference to the ability to predict personality traits. Berkovsky et al. (2019) showed 21 participants a series of affective images collected from the international affective picture system (Mikels et al., 2005) before collecting a range of personality measures via questionnaire using the HEXACO and Dark Triad conceptual frameworks (Lee & Ashton, 2014). The authors created descriptions of blink rates, fixation and saccadic behavior, and pupil size (12 metrics) in response to ten sequentially displayed stimuli with varying affective properties, totalling 120 features per participant. As such, these features describe content-based physiological behavior in response to non-competing elements. Within a classification paradigm, the authors split personality scores into three categories using equal frequency binning. After conducting feature selection using correlation-based methods, a selection of algorithms were trained to predict each of the personality traits. It was found that naive Bayes classifiers were the most performant, with the ability to classify a wide range of traits with high accuracy; specifically for Openness (85%), Conscientiousness (80.95%), Extroversion (80.95%), Agreeableness (90.48%) and resiliency (equivalent to Neuroticism under the big five theory; 80.95%).

The above research illustrates that large improvements in classification accuracy can be achieved using emotive stimuli and feature selection. However, we note that the reliance upon the HEXACO model of personality (Ashton & Lee, 2007), whilst exhibiting substantial overlap with the big five approach, does not allow direct comparison to previous literature employing the big five model of personality (Lee & Ashton, 2014). Furthermore, the reliance upon sequentially viewed static images in Berkovsky et al. (2019) is not necessarily reflective of naturalistic viewing behavior, where multiple objects in the visual scene are competing for the participant's visual attention (Tatler, Hayhoe, Land, & Ballard, 2011). As such, it remains to be seen if these findings will replicate in naturalistic environments.

### 4.2.3 Aims

It is still unknown whether above-chance classification accuracy can be achieved from descriptions of visual behavior upon social media websites. Web pages contain novel interactions, such as scrolling, which introduces partial and full occlusion of objects.This provides an environment where the user must choose whether to exploit the current content or to explore to find content which is currently out of sight. As such, this visual environment is new and distinct from previous literature using static images. Social media also provides an ecologically valid 'real world' setting that contains social and affective properties, and that is of interest for applications within human computer interface (HCI) design and advertising (Matz et al., 2017; Remsik et al., 2016). As such, we investigate whether the findings of previous literature (Azucar, Marengo, & Settanni, 2018; Berkovsky et al., 2019; Hoppe et al., 2018; Rauthmann et al., 2012; Settanni et al., 2018) generalise to this new environment. To compare the results with a range of previous literature, we choose to evaluate personality using the NEO-FFI 60 item scale under the big five model (McCrae & Costa, 2004). Furthermore, an open question remains regarding the relative contribution of different metrics describing visual behavior. Previous literature has successfully used metrics describing the physiological properties of the eye (Hoppe et al., 2018), and visual behavior in response to different types of content (Berkovsky et al., 2019). Whilst the performance of Berkovsky et al. (2019) models far outperformed Hoppe et al. (2018), several varying factors between the two studies prevent comparison.

This includes the nature of the task, the affective properties of the visual scene, the personality framework employed and the machine learning methodology followed.

Importantly, whilst it may be tempting to hypothesize that differences in visual behavior in response to emotive images are most informative of personality (Berkovsky et al., 2019), the vast differences in stimuli and methodology prevent such inferences from the current literature. To remedy this, we compare the performance of models trained upon visual behavior described in a content-agnostic manner (measuring the physiological properties of the eye) against the performance of models trained upon descriptions of content seeking behavior, in predicting trait personality from social media style websites. Finally, in a novel approach, we explore whether there is a privileged role of describing visual behavior in response to content that is distinct from location. To achieve this, we introduce five versions of a new web page. Across the five versions, we show each content at each possible location within a repeated counterbalanced design. In this way, we are able to build two separate descriptions of oculomotor behaviour across the five viewings - one set built upon descriptions of visual behavior linked to a particular location (note this is not content-specific), and the other set built upon descriptions of visual behavior in relation to a particular type of content (where the location will vary). By evaluating the performance of classifiers in predicting an individual's personality when trained upon each set, we provide insight into the unique contribution of location-specific and content-specific visual behaviour in reflecting trait-congruent visual behaviour within an ecologically valid setting.

In conclusion, in this paper we take an exploratory approach to address whether personality traits can be predicted from visual behavior in response to a social media styled website. Specifically, in the first stimulus shown we investigate whether it is best to describe the overall physiological properties of the eye over time, or employ spatial attention-based metrics. In the second stimulus shown, we investigate whether there a privileged role of content that is distinct from location, that leads to descriptions of visual behavior that reflect trait-gaze connections.

## 4.3 Methods

This study re-analyses the data collected from a previous study in chapter 3. This consists of 38 undergraduate university students ($M_{age} = 20.36$, $SD_{age} = 3.05$, 33 Female) ranging from 18-31 years old and having normal or corrected to normal vision. As in the previous study, three participants (all female) were excluded from the analysis due to having less than 80% valid samples within the eye tracking data and the Institutional Research Ethics Committee granted ethical approval. Detailed information about the stopping rule and descriptions of this cohort's viewing behaviour are available in chapter 3, although it should be noted that the power analysis is not relevant for this machine learning study.

### 4.3.1 Visual Stimuli and Procedure

Two social media inspired web pages were created, each containing five posts. Each web page was designed to emulate the Facebook platform, with user comments and company advertisements removed. However, due to practical constraints, our stimuli were not identical to a typical Facebook page. After piloting, we found that we wished to remove the ability for participants to navigate away from the page via hyperlinks (e.g., accidentally clicking and being redirected to an uncontrolled page). Secondly, we wished to ensure that the potential viewing experience was identical across participants, and to achieve this, we had to ensure that the interface would not update to a new design (e.g., during company A/B testing) during data collection. Finally, we also wished to reduce the delay between the participant's scrolling and new visual content appearing that was induced by dynamic web page loading. To address these concerns while maintaining the ability for participants to scroll down to discover new content, we created our web pages by hosting long images (e.g., as shown in Figures 4.1 and 4.2) within an HTML table. This provided a lightweight, static, and consistent base that ensured the viewing experience available was identical across participants. Having designed our stimuli, each web page was spatially divided into areas of interest (AOI) to describe oculomotor behavior according to each of the social media posts, and the other user's responses to each post (e.g. the number of likes and shares).

**Demonstration Web Page**

A social media style demonstration page was created and shown to ensure the participant was familiar with interacting with the format. Data from this interaction was not analysed and served only to ensure the participant was comfortable with the design.

**University Themed Web Page**

A university themed web page (see Figure 4.1) was created and shown once to the participant. For each of the five central image-based posts, we created AOI that represented the main content (e.g., 'Library Image') and the other users' reactions to the content.

**Animal Themed Web Page**

Five versions of an animal themed web page were created. Each version consisted of five central image-based posts, with AOI for each post and its associated reactions. The only difference across the five versions was the location of the image-based posts, with each post being shown at each of the five central locations. A single version is shown in Figure 4.2. Specifically, participants were not told that they would be viewing the same content five times to minimise the certainty of what might come next. This was because participants knowing what content is likely to come next will affect the concept of "exploring" the web page, which is undesirable.

### 4.3.2 Questionnaire Stimuli

To measure personality scores, participants completed an online version of the 60 item NEO Five-Factor Inventory (NEO-FFI; McCrae & Costa, 2004), which was presented on the Qualtrics platform. The NEO-FFI consists of five subscales (Openness to Experience, Conscientiousness, Extroversion, Agreeableness, Neuroticism) with each subscale ranging from zero (low) to 48 (high). Each subscale is represented by 12 question items that reflect a different stable facet of the individual's personality (Table 4.1).

TABLE 4.1: NEO-FFI Big Five Personality Trait Subscales (e.g., McCrae & Costa, 2004).

| Trait | Description | Example Questionnaire Item |
|---|---|---|
| Openness to Experience | High Scorers are imaginative, insightful, curious and adventurous | 'I am intrigued by the patterns I find in art and nature.' |
| Conscientiousness | High Scorers take a structured approach and enjoy a strict schedule. | 'I keep my belongings clean and neat.' |
| Extroversion | High scorers enjoy being the center of attention and meeting new people. | 'I like to have a lot of people around me.' |
| Agreeableness | High scorers are helpful and co-operative, seeking to please others. | 'I try to be courteous to everyone I meet.' |
| Neuroticism | High scorers are prone to experiencing negative affect, such as anxiety. | 'I often feel inferior to others.' |

### 4.3.3 Eye Tracking

All stimuli were presented on a 23-inch TFT monitor (1920 x 1080). The viewing distance was 60cm, and each web page was presented full-screen using the Internet Explorer browser. As the stimulus dimensions exceed one screen worth of content and participants are able to scroll down, fixations and saccades are mapped onto a single reference image using the inbuilt Tobii Studio software functionality. Eye movement data was collected using the Tobii TX300 (Tobii inc., TX300) at 300 samples per second using the Tobii Studio software. A five-point calibration procedure was followed, with visual inspection of accuracy for each participant. Fixations and saccades were detected using the Tobii Studio standard I-VT filter (Window length: 20ms, Velocity threshold: $30^{\circ}/s$). Participant data must pass an inclusion criterion of over 80% validity within the eye samples tracked.

### 4.3.4 Procedure

Participants first were briefed and gave informed consent before beginning the eye tracking segment. After verifying a successful 5-point calibration, participants viewed and scrolled through a demonstration page to ensure familiarity when interacting

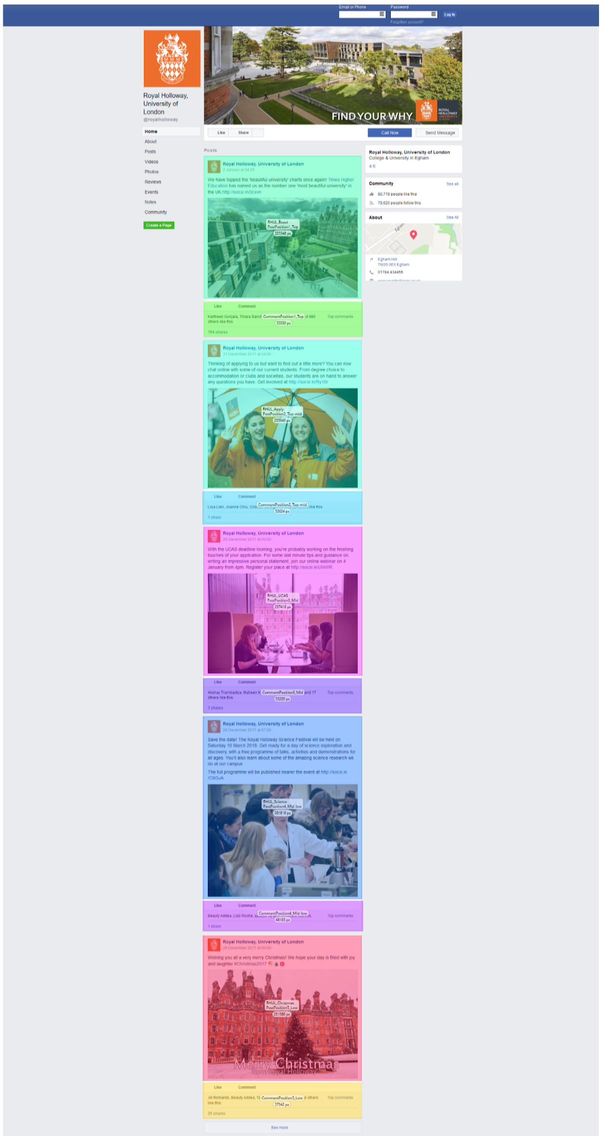

FIGURE 4.1: University themed web page stimulus. Coloured overlaid boxes and call-outs illustrate areas of interest (AOI) and were not visible to participants, with two AOI (content, other users' reactions) to each post. Each post is 13.2cm (12.55° degrees visual angle) horizontally and between 15-18cm (14.25° - 17.06°) vertically. The five central posts were evaluated for allocation of spatial attention metrics. Visual behavior across the viewing duration was evaluated for eye movement statistic metrics.

with the stimulus. After verbal confirmation of familiarity with the paradigm, written instructions appeared upon the page, instructing the participant to "view the following pages as you would normally". This was followed by a single 30-second presentation of the university Facebook page (Figure 4.1). After viewing the university web page, the participant was repeatedly presented with a version (selected

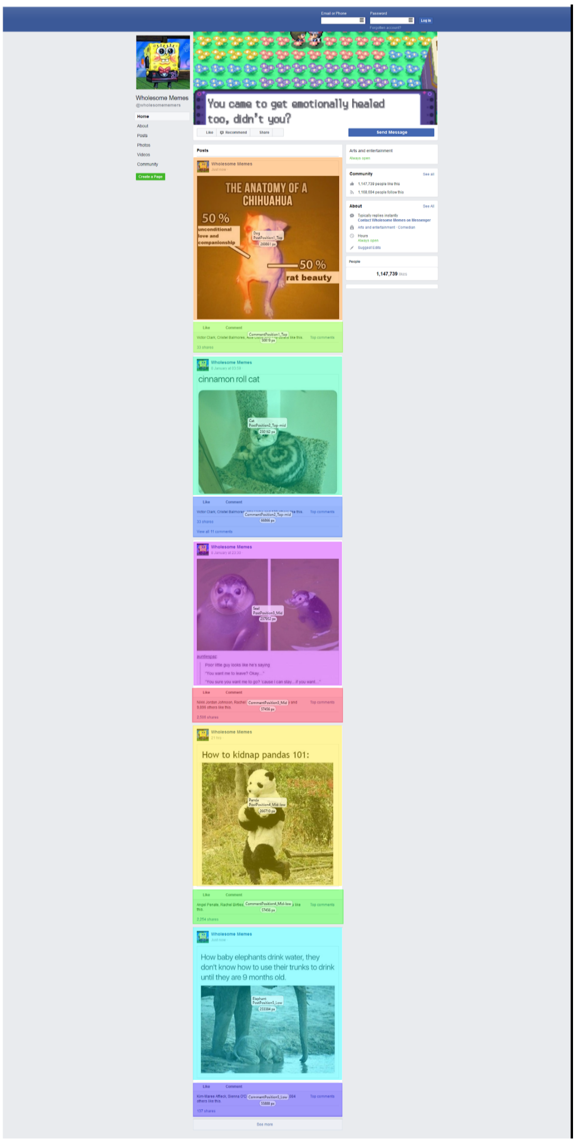


FIGURE 4.2: Animal themed web page stimulus (one of five counterbalanced versions). Coloured overlaid boxes and call-outs illustrate areas of interest (AOI) and were not visible to participants, with two AOI (content, other users' reactions) to each post. Each post is 13.2cm (12.55° degrees visual angle) horizontally and between 15-18cm (14.25° - 17.06°) vertically. The five central posts were evaluated for content-based and location-based allocation of spatial attention metrics

randomly without replacement) of the five animal web pages for thirty seconds until all of the five versions had been viewed. For example, upon first encounter the participant may see the post ('dog') at the top central position as depicted in Figure 4.2, but upon second viewing this post may be at the bottom central position. Each web page was interleaved with the presentation of a fixation cross (white screen, black cross located centrally) for five seconds. After completing eye tracking, participants completed the NEO-FFI questionnaire (McCrae & Costa, 2004). We did not counterbalance the order of the tasks (eye tracking, questionnaire) as we wished for the participants' visual behaviour to be as naturalistic as possible, and unbiased by expectations of what the experimenter may be interested in.

### 4.3.5 Machine Learning Investigation

We approach the prediction of trait personality as a supervised classification problem by splitting each of the five traits of interest into two categories (low, high) using a quantile based binning strategy. We note that our data contained multiple identical personality score values, which prevents a perfectly balanced partition into the low, high categories. Thus, for each personality trait we report the most frequently occurring category. As an example, for Openness to Experience 54.3% of the samples are in the high category. As such, an accuracy of 54.3% (rather than 50% chance) is our baseline for that trait. We build a separate model for each trait under the big five personality construct (McCrae & Costa, 2004). Following guidelines for learning from imbalanced data (He & Garcia, 2009), we also evaluate our final models using the area under the receiver operating curve (AUROC) metric (Fawcett, 2006). As we have relatively few observations this creates a high risk of overfitting, where the model performs well upon the data it has seen but fails to accurately predict new cases (Guyon & Elisseeff, 2003). As such, we take two steps to minimise this risk.

As this is an exploratory paper, there is no way to know *apriori* which algorithm will perform best for each of our classification tasks. As such, we use a range of different algorithms to predict each of the personality traits. This allows us to capture a variety of associations between the participant's oculomotor behaviour and their personality attributes. To reduce the risk of over-fitting (where the model performs well upon training data but generalises poorly to observations it has never seen before) we investigate relatively simple algorithms which have fewer degrees of freedom by which to over-fit the data. These are: k-nearest neighbours using Euclidean distance (Guo, Wang, Bell, Bi, & Greer, 2003), ridge classification (Friedman, Hastie, & Tibshirani, 2010), linear support vector machines (Hearst, Dumais, Osuna, Platt, & Scholkopf, 1998), and naive Bayes classifiers (Zhang, 2004). As these methods are sensitive to scale, we independently scale each oculomotor metric using a standardization procedure (subtracting the mean and dividing by the standard deviation). An important note here is that we choose not to directly compare the performance of the different machine learning algorithms, and simply report the best performance achieved for each trait. The rationale for this is firstly to limit the number of comparisons conducted in this study to avoid spurious results, and secondly it is because the algorithms are sensitive to the number of instances available to learn from, as well as the type of association between the input values (i.e., eye movements) and outcome variable. Given our small sample size the results from comparing the different algorithms may not be meaningful or reliable for other research groups.

Secondly, we evaluate the model's performance upon its ability to generalize to new, unseen data points, by using a nested cross-validation procedure (Outer loop: Leave-One-Out, Inner loop: Five-fold). This is essential to ensure that we avoid overestimating the performance of the classifier. For classifiers with an accuracy greater than baseline, we calculate label permutation based p-values to discover whether the classifier has identified real class structure in the data (Ojala & Garriga, 2010). If the classifier is significantly better than chance, we fix the algorithm's hyper parameters to the most frequently selected across the outer folds and calculate the area under the receiver operator curve (AUROC; Fawcett, 2006) using an additional leave one out cross-validation scheme. All analysis is conducted using the Scikit-learn python library (Pedregosa et al., 2011). Statistical values are not corrected for multiple comparisons.

### 4.3.6 Exploring visual behavior in response to the university web page

In this early exploratory stage, we focus upon the events of fixations (where the eye is relatively still) and saccades (a high-speed ballistic movement of the eye), as these provide the clearest measure of overt attention and shifts in attention (Hayhoe & Ballard, 2005). When describing fixations or saccades, we can either attribute these events to particular locations upon the stimuli, known as an area of interest (AOI) approach, or describe general statistical patterns of oculomotor behaviour that do not correspond to any particular region upon the page (Eye Movement Statistics). In this paper, we explore both options, where our AOI concern the central five images (e.g., dog, cat, etc., as illustrated in Figure 2) and their associated comments.

#### AOI-Based Features

We describe the time to first fixation (TTFF), total fixation duration (TFD), and the number of visits associated with each AOI. For TTFF, we measure how long it took from the stimulus onset (start of the web page presentation) to when the participant first makes a fixation within the AOI (time to first fixation). For TFD, we measure how long in total the participant spent fixating within the AOI. For the number of visits, we measured how many times the participant fixated inside the AOI, fixated outside the AOI, then returned to the AOI (visits). This leads to 30 features.

#### Eye Movement Statistics (EMS)

EMS properties can be considered more 'physiological' properties of how participants move their eyes in general, regardless of where or what they are looking at. We consider fixations to have one property, which is duration. We consider saccades to have two properties, duration and amplitude. We describe these three properties via the minimum, maximum, number, 25th percentile, 50th percentile (median), 75th percentile, mean, and standard deviation. As a preprocessing stage, we remove two metrics: minimum saccade duration and the number of saccadic amplitudes. We noted that the minimum duration for saccades was uniform across participants and thus held no discriminative value as a feature. Similarly, saccadic amplitudes are calculated from one fixation location to another. Due to this constraint, the number of saccade amplitudes is directly tied to the number of fixations and leads to redundant information. After removing these metrics, a total of 22 features remain.

### 4.3.7 Exploring Visual behavior upon the Animal Web Page

Within this second analysis, we explore whether there is a privileged role of content-seeking visual behaviour in reflecting personality, that is distinct from visual behaviour in response to a spatial location upon the web page. To elaborate, we try to disentangle whether personality is linked to the purposeful seeking of a particular type of content (e.g., cat, dog, etc.), or a habitual viewing pattern e.g. being biased towards the first item presented. Note that any 'purposeful seeking' would be self-directed by the participant, as we presented this as a free-viewing task. We calculate the same fixation and saccadic events as outlined within experiment one, but attribute these to either a location (e.g., the top post) or a content (e.g. 'cat' stimuli, where the spatial location will vary across presentations). This is possible by averaging each metric across the five presentations, such that the location AOI represent a variety of content but a single location, and the content AOI represent a variety of spatial locations but only one type of content. As we are also interested in the relative performance of content and location-seeking descriptions compared to EMS type description, we additionally calculate EMS for these web pages (as in experiment one) and collapse by summing each metric across the five presentations. This results in a group of features describing content-seeking behaviour (content; 30 features), location-seeking behaviour (location; 30 features), and the physiological properties of the eye movements across the five presentations (EMS; 22 features).

## 4.4 Results

### 4.4.1 Essential Descriptives

To prepare the data for a machine learning investigation, multiple manipulation checks were conducted to ensure that both the personality scores and the visual behavior metrics reflect expected properties. As such, we outline these checks below.

#### Personality Scores

Inspection of the questionnaire data revealed that, for all participants, all items from the 60-item personality questionnaire had valid responses with no missing items. The descriptive statistics and Shapiro-Wilk test for normality are presented in table 4.2. Results suggest that each personality trait is drawn from a normal distribution, which matches the expectations for the big five personality traits (McCrae & Costa, 2004). The mean and standard deviation of the categorical classes formed for each

TABLE 4.2: Descriptive statistics for personality trait scores (out of 48). Baseline accuracy represents the highest accuracy that could be achieved by classifying all instances as the dominant class (low or high) regardless of visual behavior. Independent t-tests show a significant difference ($p < .001$) between the low and high groups for each trait. The Shapiro-Wilk test for normality does not suggest any traits are non-normally distributed ($p > 0.188$).

| Label | Overall Score Mean (SD) | Low Scorers Mean (SD) | High Scorers Mean (SD) | Baseline % (class) |
|---|---|---|---|---|
| Openness | 30.11 (6.12) | 24.75 (3.34) | 34.63 (3.62) | 54.3 (High) |
| Conscientiousness | 30.11 (8.50) | 22.88 (4.48) | 36.94 (4.64) | 51.4 (High) |
| Extroversion | 30.54 (6.52) | 25.18 (4.30) | 35.61 (3.15) | 51.4 (High) |
| Agreeableness | 31.20 (6.19) | 26.35 (3.98) | 35.78 (3.78) | 51.4 (High) |
| Neuroticism | 24.51 (7.13) | 18.50 (2.98) | 29.58 (5.24) | 54.3 (High) |

personality trait are also shown in Table 4.2, where we also test to establish that there is a significant difference between the two groups (low, high) for each trait.

### 4.4.2 Individual Variances in Visual Behavior

To understand the visual behaviour being evoked by our novel stimuli across participants, we report descriptive statistics for the time to first fixation (TTFF) and total fixation duration (TFD) metrics upon the central five AOI in each of our web pages.

#### University Web Page

We observe that the TTFF is influenced by the position of the content upon the page. For example, on average, participants fixate upon the top post sooner than the bottom post (Figure 4.3A) with a general trend of decreasing median total fixation duration as the participant progresses from the top to the bottom of the web page (Figure 4.3B). This is likely due to the time constraint imposed upon the participants, who often reached the bottom of the web page towards the end of the trial. Regardless, this supports that our participants had enough time to reach each AOI and exhibit individual variation in total fixation duration upon each content.

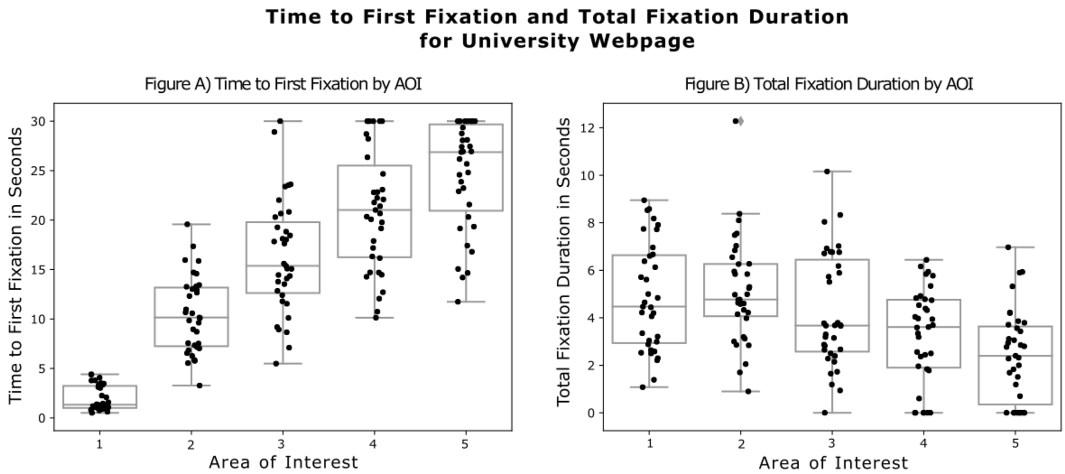


FIGURE 4.3: Box plot illustrating time to first fixation (A) and total fixation duration (B) upon the five central content areas of interest within the University web page. Median line and interquartile range are shown, with whiskers permitted 1.5 times the interquartile range. Regions are ordered by their location upon the page, from top (1) to bottom (5).

**Animal Web Page**

Visual metrics derived from the animal web page represent visual behavior averaged across five presentations. This allowed the encoding of AOI by both content location (e.g., top, bottom) and by content type (e.g., cat, dog). Here we check whether we have successfully decoupled the association between content and location. In the ideal situation, there would be only slight variations in the time to first fixation between our AOI when encoding by content type after averaging across the counterbalanced versions. Visualising the time to first fixation when AOI are encoded by content (Figure 4.4A) compared to when the AOI are encoded by location (Figure 4.4B) supports the idea that the counterbalancing procedure has successfully decoupled the association between content and location.

### 4.4.3 Main Results

Here, we address the questions of interest to this study. Namely, we look at the performance of classifiers trained to predict the personality traits of interest using different visual metrics as input. We first cover results from the university page, asking whether overall statistical properties or attention-based metrics best reflect trait-gaze

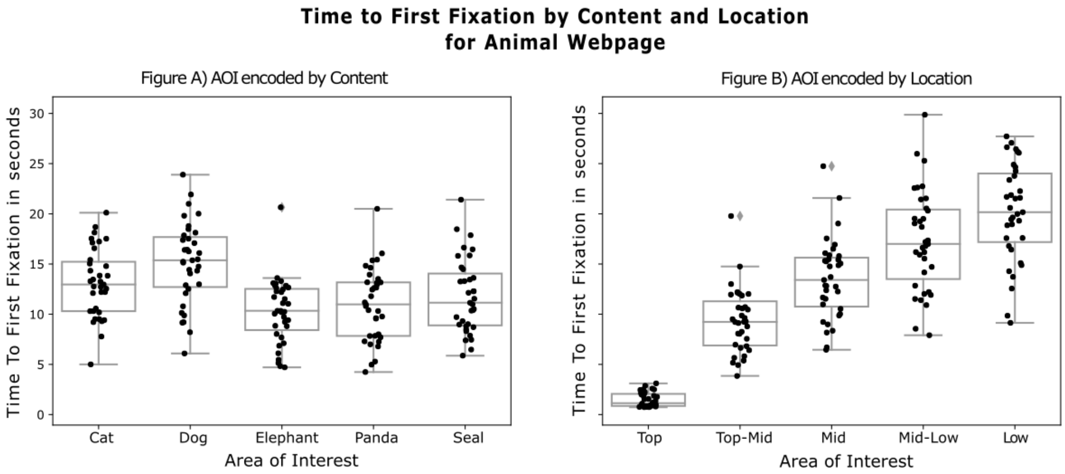


FIGURE 4.4: Box plot illustrating time to first fixation upon the five central content (A) and location (B) areas of interest within the Animal web page. Median and interquartile range are shown with whiskers permitted 1.5 times the interquartile range. Regions are ordered by their location upon the page, from top to bottom ('low'). The y-axis is shared.

associations, before moving on to investigate whether content-seeking and location-based visual behavior have distinct roles in reflecting trait-gaze associations.

**University Web Page**

By applying our machine learning framework to predict the traits of interest from visual behavior upon the university web page, we address whether it is best to describe the overall statistical properties of eye movements over time, or spatial attention-based metrics (EMS vs ASA features). Our results show that spatial attention based metrics of oculomotor behavior (ASA features) can be used to predict the trait of Agreeableness ($AUROC = .87$) significantly better than chance using a Ridge classifier. While the remaining traits (except for neuroticism) were able to be predicted with an accuracy above that expected by chance (baseline), the models were not significantly better than chance (See table 4.3).

TABLE 4.3: University Web Page Model Performance. Best model and feature group combinations for each personality trait.P-value: label permutation based p-value calculated across 200,000 runs. k-NN: k-nearest neighbours algorithm. SVM: linear support vector machine algorithm. Ridge: Ridge Classification in one vs all scheme.

| Label | Algorithm | Feature Group | Accuracy | P-Value |
|---|---|---|---|---|
| Openness | SVM | EMS | 66% | .067 |
| Conscientiousness | Ridge | ASA | 60% | .200 |
| Extroversion | Naive Bayes | EMS | 57% | .200 |
| Agreeableness | Ridge | ASA | 71% | .013 |
| Neuroticism | Naive Bayes | EMS | 54% | .450 |

**Animal Themed Web Page**

We also applied our machine learning framework to predict personality trait categories using visual behavior across the five versions of the animal web page. This analysis evaluates whether content-based oculomotor behaviour (e.g., the participant displayed this behaviour whilst viewing the 'cat' based post), location-based oculomotor behaviour (e.g., the participant displayed this behaviour whilst viewing the top post) or statistical descriptions of visual behaviour (which are not location or content bound), lead to the best performance when predicting personality categories from visual behaviour. We find classifiers trained upon the statistical (EMS) features are able to predict the traits of Openness to Experience (Ridge: $AUROC = .80$, SVM: $AUROC .84$), Extroversion ($AUROC = .76$) and Neuroticism ($AUROC = .74$) significantly better than chance (see table 4.4). In contrast, only the trait of Openness ($AUROC = .72$) was able to be predicted significantly better than chance using the location-based descriptions. Finally, whilst many classifiers trained upon content-based features performed above baseline, none performed significantly better than chance.

TABLE 4.4: Animal Web Page Model Performance. Model and feature group combinations that perform above baseline for each personality trait. P-value: label permutation based p-value calculated across 200,000 runs. k-NN: k-nearest neighbours. SVM: linear support vector machine algorithm. Ridge: Ridge Classification in one vs all scheme.

| Label | Algorithm | Feature Group | Accuracy | P- Value |
|---|---|---|---|---|
| Openness | Ridge | EMS | 74% | .002 |
| | SVM | EMS | 71% | .014 |
| | SVM | Location | 69% | .014 |
| | KNN | Location | 63% | .067 |
| | Ridge | Location | 63% | .067 |
| Conscientiousness | KNN | Content | 60% | .201 |
| | Naive Bayes | Content | 60% | .201 |
| Extroversion | Ridge | EMS | 71% | .013 |
| | SVM | EMS | 66% | .063 |
| | Naive Bayes | EMS | 66% | .063 |
| | Naive Bayes | EMS | 57% | .200 |
| Agreeableness | Ridge | Location | 54% | .434 |
| | SVM | Location | 54% | .434 |
| Neuroticism | KNN | EMS | 69% | .015 |
| | KNN | Content | 66% | .068 |
| | SVM | EMS | 66% | .068 |
| | Ridge | EMS | 63% | .068 |
| | Ridge | Location | 57% | .210 |

**Exploratory Analysis**

Having established the relative performance of each feature group when predicting personality traits from visual behaviour, we evaluate the best performance achievable by combining the feature groups for each experiment. Due to this resulting in a minimum of 52 features, we employ principle component analysis as a dimensionality reduction technique. For both experiments, a single principle component is used to capture over 95% of the variance across the features. We report results that exceed the performance of the classifiers presented above. For experiment one, the combined feature group (ASA and EMS) outperforms all others for the traits of Extroversion ($77\%_{accuracy}, p = .002, AUROC : .77$) using a nearest neighbour algorithm, and Neuroticism ($60\%_{Accuracy}, p = .210$) using a Ridge algorithm. For experiment two, the combined feature group (EMS, Content, Location) outperforms all individual feature groups for the trait of Agreeableness ($60\%_{accuracy}, p = .201$) using a k-nearest neighbours algorithm.

## 4.5 Discussion

In a data-driven approach, we have addressed the question of whether visual behavior in response to a social media style website is informative of personality. This novel visual environment provides web-based user interactions such as scrolling, and contains visual stimuli with both social and affective properties. This clearly distinguishes the properties of this environment from previous research using static images (Berkovsky et al., 2019), abstract animations (Rauthmann et al., 2012), and real-world locomotion conditions (Hoppe et al., 2018).

Our findings demonstrate that the eye movements of our participants were informative of Openness to Experience, Extroversion and Neuroticism when they viewed the animal themed web page, and Agreeableness when viewing the university themed web page. Previous literature has identified that trait personality is reflected in a user's likes (Kosinski et al., 2013; Youyou, Kosinski, & Stillwell, 2015), linguistic features (Tandera, Hendro, Suhartono, Wongso, & Prasetio, 2017) and other online behaviours (Azucar et al., 2018) whilst engaging with social media content. Our findings complement this literature by illustrating that machine learning algorithms can predict specific personality traits significantly better than chance from an individual's visual behavior upon a social media style website. Additionally, we identify that, within our paradigm, above-chance personality trait prediction can be achieved after collecting between thirty seconds (university themed web page) and two and a half minutes (animal themed web page) of eye tracking data. This stands in contrast to previous literature, such as Hoppe et al. (2018) who collected data over $\sim 10$

minutes. Thus, we illustrate that above-chance predictions of personal traits can be established from eye movements over a much shorter period than previous known.

The ability to deploy eye-tracking based personality detection rapidly improves the usefulness of the technique as a tool for psychological research (Kosinski et al., 2015), and has direct applications in informing the personalizing of online displays (Al-Samarraie, Eldenfria, & Dawoud, 2016) and advertisements (Matz & Netzer, 2017) to suit the individual's personality. Specifically, the traits of Openness to Experience and Extroversion (predicted significantly better than chance from the animal themed web page) have been found to be useful for persuasive appeals on social media site platforms, with Kosinski and Wang (2017) finding that matching advertisements to suit the viewer's score upon one of these traits (i.e., whether they are high or low) resulted in up to 40% more clicks and up to 50% more purchases than showing a mismatching or unpersonalized advertisement. As individuals with high neuroticism are more likely to compulsively buy fashion clothing items (Johnson & Attmann, 2009), we speculate that the group may be desirable to target with so called 'fast fashion' items (i.e., low cost replications of current catwalk trends and high-fashion designs). However, the same high neuroticism group is more likely to have negative emotional experiences (Costa, 1996), and as such we speculate that they may be undesirable to target when requesting product reviews. The above examples are by no means exhaustive, but serve to illustrate the wide-ranging potential impact of being able to quickly evaluate and understand the psychological attributes of a given individual.

Additionally, this work establishes that statistical patterns that generalise across participants can be detected from visual behavior within a free-viewing design. Within free-viewing designs, due to a lack of a directed task, participants may choose to interpret the task as they will. As visual behavior is heavily influenced by task (Tatler et al., 2010), across participants, these variances in task interpretation can act to introduce individual variation in oculomotor behavior. This can be conceptualised as introducing random variance into the visual data, or as an additional degree of freedom that allows participants the opportunity to express trait-congruent behavior. To reduce random variance, it is advisable to introduce a directed task to make the visual behavior across participants directly comparable. However, this reduces the ecological validity within our design, and our results support that for the traits of Openness to Experience, Extroversion and Neuroticism, a directed task is not necessary. Future research may wish to investigate whether stable trait-gaze associations emerge within a range of different tasks, such as within visual search or memory tasks (Poynter, Barber, Inman, & Wiggins, 2013).

Finally, as personality profiling techniques have found to be effective in driving consumer loyalty (Matz et al., 2017) and consumer behavior via social media sites (Matz & Netzer, 2017), our findings have implications upon privacy policy. Many online users express negative reactions to targeted advertising (Purcell et al., 2012), and it is currently unreasonable to expect the public to understand that by providing their eye movements they may be unknowingly disclosing implicit information about their personal traits and attributes. As such, whilst there is potential for personality profiling techniques to promote positive actions, such as environmentally friendly behaviours, the covert and invasive nature of these methodologies in persuading an individual's actions and behavior without the user's knowledge or consent is troubling. We suggest that future research investigates the public perception and opinion upon the use of visual behavior to covertly predict personality-trait scores within SNS settings, including whether individuals generally consider this an invasion of privacy. This point is especially salient in light of data privacy concerns surrounding social media site use (Kosinski et al., 2015), and the findings that Facebook has intentionally and knowingly violated both data privacy and anticompetition laws (Committee, 2019).

Having established that it is possible to predict the personality traits of Openness to Experience, Extroversion, Agreeableness and Neuroticism significantly better than chance within our paradigm, we direct our attention to the proposition that attention-based descriptions of oculomotor behavior in response to different types of content within the visual scene would provide useful information for the prediction of these personality traits from visual behavior. We also address the question of whether visual behavior in response to content has a unique role in eliciting trait-gaze behavior, that is more informative of personality than visual behavior in response to a location upon the page.

### 4.5.1 Allocation of Spatial Attention vs Eye Movement Statistics

Within our second analysis concerning the animal-themed web page, we find that statistical descriptions of oculomotor behavior ('EMS") produce classifiers capable of predicting the traits Openness, Extroversion and Neuroticism significantly above chance. We note that, whilst comparatively weaker in performance than the EMS-based metrics, that models trained upon visual behaviour in response to a location upon the web page can also be used to predict trait openness to experience significantly above chance. As such, within short single presentation designs, we suggest the most useful descriptions of oculomotor behavior for personality trait prediction are those describing the statistical properties of fixations and saccades across the viewing duration, rather than the allocation of visual attention to a particular region. This implies that general eye movement characteristics across the page capture useful individual variance that is informative of Openness to Experience, Extroversion and Neuroticism. This may inform research aimed at creating socially aware HCI interfaces (Al-Samarraie et al., 2016) within social media websites, by supporting the integration of features describing oculomotor event parameters across the page rather than in response to spatial regions. Furthermore, our findings contradict Berkovsky et al. (2019) in that we do not identify naive Bayes to be the optimal classifier for predicting personality from oculomotor behavior. Instead, we identify that k-NN, Ridge and linear support vector machine algorithms often outperform naive Bayes when applied to our feature sets. This is likely due to key differences in the nature of our visual stimuli (interactive web page vs static images) and methodology, which reflect diverging aims. Whereas we wished to understand the predictive power of two distinct ways of describing visual behavior, Berkovsky et al. (2019) were interested in finding a minimal set of features (regardless of derivation) that built a strong classifier. To achieve this, they implemented correlation based feature selection, which seeks to minimise inter-feature correlations whilst maximising feature-label correlations (Hall & Lloyd, 1999). This reduction of inter-feature correlation directly mirrors the naive Bayes assumption that features are independent from each other (e.g., have little or no covariance). Due to the inter-correlated nature of the variables in our data (e.g., number of fixations and saccades), we violate this independence assumption, which disadvantages a naive Bayes classifier and may explain the difference in findings. As such, it appears that naive Bayes is not the optimal classifier for classifying personality traits from visual data when comparing sets of features without correlation based feature selection. Based upon our results, in similar paradigms we would suggest that, where possible, the relationship between personality and visual behavior should be captured using linear methods such as Ridge Classification, that retain the interpretability of Bayesian methods and are not heavily reliant upon the independence of features assumption. When not possible, k-nearest neighbors appear to provide a viable non-linear alternative. However, we note that this is based upon results from a small sample, and this may not generalise to cases where there are a greater number of observations to learn from.

### 4.5.2 Does visual behavior in response to content have a unique role in eliciting trait congruent gaze behavior?

Our findings do not provide support for the existence of a privileged role of visual behavior in response to content in eliciting trait congruent gaze behavior. While this may suggest that the findings of previous literature (Berkovsky et al., 2019; Rauthmann et al., 2012) do not replicate within the ecological niche of social-media based web page stimuli, a number of alternative explanations must be addressed first. For example, our areas of interest were inspired by previous literature illustrating that 'liking' behavior (Kosinski et al., 2013) and visual responses to content with social and affective properties, (Berkovsky et al., 2019; Nummenmaa et al., 2006) were informative of a range of personal attributes. As such, we created areas of interest that highlighted the main post content (e.g., the image) and the reactions to that content. However, many alternative labelling strategies exist, and it is possible that our scheme underestimates the influence of personality upon the expression of content seeking visual behaviour. Furthermore, the tight layout of the Facebook style social media site leaves little to no boundary between each of the elements within a single post (content, comments). In our design, this forced neighboring areas of interest which border upon each other. Due to inaccuracies in both the human visual system and our eye tracking system (which allows free movement of head), this may have introduced statistical noise via the attribution of false positives/negatives across participants (Orquin & Holmqvist, 2018). Future literature may wish to explore a range of different AOI coding schemes over a more diverse range of stimuli to discover if the performance of classifiers built upon AOI-based features varies with different encoding strategies.

### 4.5.3 Caveats

Importantly, there are caveats to our research which must be acknowledged. Firstly, our analysis took place upon a university student cohort of thirty-five participants. This provides a relatively small sample from which to generalise from, and future research will be required to establish whether these findings generalise to more diverse cohorts. Furthermore, to allow for direct comparison across participants, in analysis one we displayed an identical web page to all participants. This allowed us to directly compare the usefulness of describing eye movements via statistical vs AOI-based metrics, but does not directly marry to the more ecologically valid setting of each individual having a personalized news feed tailored to their interests and past browsing history. A direction for future research may be investigating whether individual differences in content obscure the statistical trends across participants that allow machine learning algorithms to accurately delineate low scorers from high. Furthermore, as visual routines are heavily influenced by task (Tatler et al., 2010) and visual research conducted within laboratory settings does not always replicate in ecologically valid environments (Macdonald & Tatler, 2018), it is important to understand whether our instructions to "view the following pages as you would normally" lead to visual behavior comparable to naturalistic social media use. This raises the concern of whether the models generated within this study will generalise to naturalistic viewing. As being watched is associated with a slower and more methodical visual search behavior (Miyazaki, 2013), a possible solution for future studies is to employ eye tracking within an online experimental setting. This allows the user to participate within a naturalistic setting, encouraging naturalistic social media use. This may provide a partial explanation as to why traits such as conscientiousness were not predictable from visual behaviour upon either the university or animal web page. High Conscientiousness has been linked to a thorough approach to task completion (Claessens, Eerde, Rutte, & Roe, 2010) and low Conscientiousness is linked to low task adherence (Molloy, O'Carroll, & Ferguson, 2014). If a sense of being watched encouraged all participants to employ a more principled (slower and more methodical) approach to viewing the content, this would likely mask differences in visual behaviour that may otherwise be useful for distinguishing high-conscientiousness from low in our cohort.

### 4.5.4 Conclusion

In summary, we bring together previous research that has identified social media usage (Azucar et al., 2018; Gosling, Augustine, Vazire, Holtzman, & Gaddis, 2011; Kosinski et al., 2013; Youyou et al., 2015) and eye movements (Baranes, Oudeyer, & Gottlieb, 2015; Hoppe et al., 2018; Risko, Anderson, Lanthier, & Kingstone, 2012) as two separate elements that correlate with personality trait scores. In a screen-based eye tracking paradigm, we find partial support for our hypothesis that personality traits are reflected in eye movements made while viewing a social-media based web page. We found the traits of Openness to Experience, Extroversion and Neuroticism were predicted significantly above chance, while Conscientiousness and Agreeableness were not. Furthermore, we found that our classifiers were able to perform better when trained upon statistical descriptions of eye movements compared to when trained upon AOI-based metrics.

We raise concern over the privacy issues that surround the use of such techniques in targeted advertisement settings, and apply our findings to the development of socially aware HCI. Namely, we suggest that eye movements are a vital source of social information that can be utilised in an automated manner to give insight into an individual's personality over a very short time period, thus further promoting the integration of eye-based measures into socially aware HCI interfaces (Bargary et al., 2017; Blanton et al., 2009). We suggest that future research investigates whether such findings replicate within unsupervised, non-laboratory settings, and across different visual tasks. We also suggest investigation into the public perception of using covert methods to deliver personality-targeted advertisements, including whether individuals generally consider this an invasion of privacy or a desirable feature.

## 4.6 Acknowledgement(s)

We thank Jenna Morshead for providing invaluable assistance in collecting the data.

## 4.7 Disclosure statement

We declare that none of the authors have competing interests.

## 4.8 Funding

This work was supported by a Leverhulme Magna Carta Doctoral Scholarship.

## References

Adjerid, I., & Kelley, K. (2018). Big data in psychology: A framework for research advancement. *American Psychologist*, *73*(7), 899–917. doi:10.1037/amp0000190

Al-Samarraie, H., Eldenfria, A., & Dawoud, H. (2016). The impact of personality traits on users' information-seeking behavior. *Inf. Process. Manag. 53*(1), –. doi:http://dx.doi.org/10.1016/j.ipm.2016.08.004

Armstrong, T., & Olatunji, B. O. (2012). Eye tracking of attention in the affective disorders: A meta-analytic review and synthesis. *Clin. Psychol. Rev. 32*(8), 704–723. doi:10.1016/j.cpr.2012.09.004

Ashton, M. C., & Lee, K. (2007). Empirical, Theoretical, and Practical Advantages of the HEXACO Model of Personality Structure. *Personality and Social Psychology Review*, *11*(2), 150–166. doi:10.1177/1088868306294907

Azucar, D., Marengo, D., & Settanni, M. (2018). Predicting the Big 5 personality traits from digital footprints on social media: A meta-analysis. *Pers. Individ. Dif. 124*, 150–159. doi:10.1016/j.paid.2017.12.018

Baranes, A., Oudeyer, P. Y., & Gottlieb, J. (2015). Eye movements reveal epistemic curiosity in human observers. *Vision Res. 117*, 81–90. doi:10.1016/j.visres.2015.10.009

Bargary, G., Bosten, J. M., Goodbourn, P. T., Lawrance-Owen, A. J., Hogg, R. E., & Mollon, J. (2017). Individual differences in human eye movements: An oculomotor signature? *Vision Res. 141*, 157–169. doi:10.1016/j.visres.2017.03.001

Berkovsky, S., Taib, R., Koprinska, I., Wang, E., Zeng, Y., Li, J., & Kleitman, S. (2019). Detecting Personality Traits Using Eye-Tracking Data. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (pp. 1–12). CHI '19. doi:10.1145/3290605.3300451

Blanton, M., Zhou, J., Moro, M. M., Zhang, D., Tsotras, V. J., Halkidi, M., … Tompa, F. W. (2009). Human-Computer Interaction. In *Encycl. Database Syst.* (pp. 1327–1331). doi:10.1007/978-0-387-39940-9_192

Bleidorn, W., & Hopwood, C. J. (2018). Using Machine Learning to Advance Personality Assessment and Theory. *Personality and Social Psychology Review*, *23*(2), 190–203. doi:10.1177/1088868318772990

Bleidorn, W., Hopwood, C. J., & Wright, A. G. (2017). *Using big data to advance personality theory*. doi:10.1016/j.cobeha.2017.08.004

Chelazzi, L., Perlato, A., Santandrea, E., & Della Libera, C. (2013). Rewards teach visual selective attention. *Vision Res. 85*, 58–62. doi:10.1016/j.visres.2012.12.005

Claessens, B. J. C., Eerde, W. V., Rutte, C. G., & Roe, R. A. (2010). Things to Do Today . . . : A Daily Diary Study on Task Completion at Work. *Applied Psychology*, *59*(2), 273–295. doi:10.1111/j.1464-0597.2009.00390.x

Committee, C. S. (2019). *Disinformation and 'fake news': Final Report*. The Digital, Culture, Media and Sport Committee.

Costa, P. T. (1996). Work and personality: Use of the NEO-PI-R in industrial/organisational psychology. *Appl. Psychol. 45*(3), 225–241. doi:10 . 1111 / j . 1464 - 0597 . 1996 . tb00766.x

Eckstein, M., Guerra-Carrillo, B., Miller Singley, A., & Bunge, S. (2017). Beyond eye gaze: What else can eyetracking reveal about cognition and cognitive development? *Dev. Cogn. Neurosci. 25*, 69–91. doi:10.1016/j.dcn.2016.11.001

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*. ROC Analysis in Pattern Recognition, *27*(8), 861–874. doi:10.1016/j.patrec.2005.10.010

Fox, E., Russo, R., & Dutton, K. (2002). Attentional bias for threat: Evidence for delayed disengagement from emotional faces. *Cogn. Emot. 16*(3), 355–379. doi:10.1080/02699930143000527

Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of statistical software*, *33*(1), 1–22.

Gosling, S. D., Augustine, A. A., Vazire, S., Holtzman, N., & Gaddis, S. (2011). Manifestations of Personality in Online Social Networks: Self-Reported Facebook-Related Behaviors and Observable Profile Information. *Cyberpsychology, Behav. Soc. Netw. 14*(9), 483–488. doi:10.1089/cyber.2010.0087

Guo, G., Wang, H., Bell, D., Bi, Y., & Greer, K. (2003). KNN Model-Based Approach in Classification. In *On the Move to Meaningful Internet Systems* (pp. 986–996). doi:10.1007/978-3-540-39964-3_62

Guyon, I., & Elisseeff, A. (2003). An Introduction to Variable and Feature Selection. *J. Mach. Learn. Res. 3*(3), 1157–1182. doi:10.1016/j.aca.2011.07.027

Hall, M., & Lloyd, S. (1999). Feature Selection For Machine Learning: Comparing a Correlation-based Filter Approach to the Wrapper. In *Twelfth International FLAIRS Conference*.

Hayhoe, M., & Ballard, D. (2005). Eye movements in natural behavior. *Trends Cogn. Sci. 9*(4), 188–194. doi:10.1016/j.tics.2005.02.009

He, H., & Garcia, E. A. (2009). Learning from Imbalanced Data. *IEEE Transactions on Knowledge and Data Engineering*, *21*(9), 1263–1284. doi:10.1109/TKDE.2008.239

Hearst, M. A., Dumais, S. T., Osuna, E., Platt, J., & Scholkopf, B. (1998). Support vector machines. *IEEE Intelligent Systems and their Applications*, *13*(4), 18–28. doi:10.1109/5254.708428

Hoppe, S., Loetscher, T., Morey, S. A., & Bulling, A. (2018). Eye Movements During Everyday Behavior Predict Personality Traits. *Front. Hum. Neurosci. 12*, 105. doi:10.3389/FNHUM.2018.00105

Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nat. Rev. Neurosci. 2*(3), 194–203. doi:10.1038/35058500

Johnson, T., & Attmann, J. (2009). Compulsive buying in a product specific context: Clothing. *Journal of Fashion Marketing and Management: An International Journal*, *13*(3), 394–405. doi:10.1108/13612020910974519

Kosinski, M., & Wang, W. (2017). Deep Neural Networks Are More Accurate Than Humans at Detecting Sexual Orientation From Facial Images | Stanford Graduate School of Business. *J. Personal. Soc. Psychol. (in Press. 114*(2), 246–257.

Kosinski, M., Stillwell, D., & Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. *Proc. Natl. Acad. Sci. 110*(15), 5802–5805. doi:10.1073/pnas.1218772110

Kosinski, M., Matz, S. C., Gosling, S. D., Popov, V., & Stillwell, D. (2015). Facebook as a research tool for the social sciences: Opportunities, challenges, ethical considerations, and practical guidelines. *Am. Psychol. 70*(6), 543–556. doi:10.1037/a0039210

Kruijff, E., Swan Ii, J. E., Feiner, S., Swan, J. E., & Feiner, S. (2010). Perceptual issues in augmented reality revisited. In *9th IEEE Int. Symp. Mix. Augment. Real. 2010 Sci. Technol. ISMAR 2010 - Proc.* (pp. 3–12). doi:10.1109/ISMAR.2010.5643530

Lee, K., & Ashton, M. C. (2014). The Dark Triad, the Big Five, and the HEXACO model. *Pers. Individ. Dif. 67*, 2–5. doi:10.1016/j.paid.2014.01.048

Macdonald, R. G., & Tatler, B. W. (2018). Gaze in a real-world social interaction: A dual eye-tracking study. *Q. J. Exp. Psychol.* 174702181773922. doi:10.1177/1747021817739221

Matz, S. C., Kosinski, M., Nave, G., & Stillwell, D. J. (2017). Psychological targeting as an effective approach to digital mass persuasion. *Proceedings of the National Academy of Sciences*, *114*(48), 12714–12719. doi:10.1073/pnas.1710966114. eprint: https://www.pnas.org/content/114/48/12714.full.pdf

Matz, S. C., & Netzer, O. (2017). Using Big Data as a window into consumers' psychology. *Current Opinion in Behavioral Sciences*, *18*, 7–12. doi:10.1016/j.cobeha.2017.05.009

Matzler, K., Bidmon, S., & Grabner-Kräuter, S. (2006). Individual determinants of brand affect: The role of the personality traits of extraversion and openness to experience. *J. Prod. Brand Manag. 15*(7), 427–434. doi:10.1108/10610420610712801

McCrae, R., & Costa, P. (2004). A contemplated revision of the NEO Five-Factor Inventory. *Pers. Individ. Dif. 36*(3), 587–596. doi:10.1016/S0191-8869(03)00118-1

McDougal, D. H., & Gamlin, P. D. (2015). Autonomic control of the eye. *Comprehensive Physiology*, *5*(1), 439–473. doi:10.1002/cphy.c140014

Mikels, J. A., Fredrickson, B. L., Larkin, G. R., Lindberg, C. M., Maglio, S. J., & Reuter-Lorenz, P. A. (2005). Emotional category data on images from the international affective picture system. *Behavior Research Methods*, *37*(4), 626–630. doi:10.3758/BF03192732

Mischel, W., & Shoda, Y. (2008). Handbook of Personality: Theory and Research. (pp. 208–241). Guilford Press.

Miyazaki, Y. (2013). Increasing Visual Search Accuracy by Being Watched. *PLOS ONE*, *8*(1), e53500. doi:10.1371/journal.pone.0053500

Molloy, G. J., O'Carroll, R. E., & Ferguson, E. (2014). Conscientiousness and Medication Adherence: A Meta-analysis. *Annals of Behavioral Medicine*, *47*(1), 92–101. doi:10.1007/s12160-013-9524-4

Nadkarni, A., & Hofmann, S. G. (2012). Why do people use Facebook? *Pers. Individ. Dif. 52*(3), 243–249. doi:10.1016/J.PAID.2011.11.007

Nummenmaa, L., Hyönä, J., & Calvo, M. G. (2006). Eye movement assessment of selective attentional capture by emotional pictures. *Emotion*, *6*(2), 257–268. doi:10.1037/1528-3542.6.2.257

Ojala, M., & Garriga, G. C. (2010). Permutation Tests for Studying Classifier Performance. *J. Mach. Learn. Res. 11*(Jun), 1833–1863.

Orquin, J. L., & Holmqvist, K. (2018). Threats to the validity of eye-movement research in psychology. *Behavior Research Methods*, *50*(4), 1645–1656. doi:10.3758/s13428-017-0998-z

Papoutsaki, A., Laskey, J., & Huang, J. (2017). SearchGazer: Webcam Eye Tracking for Remote Studies of Web Search. In *Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval* (pp. 17–26). CHIIR '17. doi:10.1145/3020165.3020170

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res. 12*(Oct), 2825–2830.

Poynter, W., Barber, M., Inman, J., & Wiggins, C. (2013). Individuals exhibit idiosyncratic eye-movement behavior profiles across tasks. *Vision Res. 89*, 32–38. doi:10.1016/j.visres.2013.07.002

Purcell, K., Brenner, J., & Rainie, L. (2012). *Search Engine Use 2012*. Pew Research Center.

Rauthmann, J. F., Seubert, C. T., Sachse, P., & Furtner, M. R. (2012). Eyes as windows to the soul: Gazing behavior is related to personality. *J. Res. Pers. 46*(2), 147–156. doi:10.1016/j.jrp.2011.12.010

Remsik, A., Young, B., Vermilyea, R., Kiekhoefer, L., Abrams, J., Evander Elmore, S., ... Prabhakaran, V. (2016). *A review of the progression and future implications of brain-computer interface therapies for restoration of distal upper extremity motor function after stroke*. doi:10.1080/17434440.2016.1174572

Risko, E. F., Anderson, N. C., Lanthier, S., & Kingstone, A. (2012). Curious eyes: Individual differences in personality predict eye movement behavior in scene-viewing. *Cognition*, *122*(1), 86–90. doi:10.1016/J.COGNITION.2011.08.014

Saalmann, Y. B., Pigarev, I. N., & Vidyasagar, T. R. (2007). Neural Mechanisms of Visual Attention: How Top-Down Feedback Highlights Relevant Locations. *Science*, *316*(5831), 1612–1615. doi:10.1126/science.1139140

Settanni, M., Azucar, D., & Marengo, D. (2018). Predicting Individual Characteristics from Digital Traces on Social Media: A Meta-Analysis. *Cyberpsychology, Behav. Soc. Netw. 21*(4), 217–228. doi:10.1089/cyber.2017.0384

Shimojo, S., Simion, C., Shimojo, E., & Scheier, C. (2003). Gaze bias both reflects and influences preference. *Nat. Neurosci. 6*(12), 1317–1322. doi:10.1038/nn1150

Stanney, K. M., Mourant, R. R., & Kennedy, R. S. (1998). Human factors issues in virtual environments: A review of the literature. *Presence Teleoperators Virtual Environ. 7*(4), 327–351. doi:10.1162/105474698565767

Tandera, T., Hendro, Suhartono, D., Wongso, R., & Prasetio, Y. L. (2017). Personality Prediction System from Facebook Users. In *Procedia Comput. Sci.* (Vol. 116, pp. 604–611). doi:10.1016/j.procs.2017.10.016

Tatler, B., Hayhoe, M., Land, M., & Ballard, D. (2011). Eye guidance in natural vision: Reinterpreting salience. *J. Vis. 11*(5), 5–5. doi:10.1167/11.5.5

Tatler, B. W., Wade, N. J., Kwan, H., Findlay, J. M., & Velichkovsky, B. M. (2010). Yarbus, eye movements, and vision. *Iperception. 1*(1), 7–27. doi:10.1068/i0382

Weaver, S. D., & Gahegan, M. (2007). Constructing, Visualizing, and Analyzing a Digital Footprint*. *Geographical Review*, *97*(3), 324–350. doi:10.1111/j.1931-0846.2007.tb00509.x

Youyou, W., Kosinski, M., & Stillwell, D. (2015). Computer-based personality judgments are more accurate than those made by humans. *Proc. Natl. Acad. Sci. 112*(4), 1036–1040. doi:10.1073/pnas.1418680112

Zhang, H. (2004). The Optimality of Naive Bayes. *AAI*, *1*(2), 6.

# Chapter 5

# Twenty Seconds to Know You? Classifying Personality from Visual Behaviour on Social Media

## 5.1 Abstract

Eye tracking allows the researcher to capture individual differences in the expression of their visual exploration behaviour, which in certain contexts has been found to reflect aspects of the user's preferences and personality. In a novel approach, we recorded the eye movements of 180 participants while they browsed their Facebook News Feed and employed a machine learning approach to predict each of the self-reported Big Five personality traits from this viewing behaviour. We identify that an individual's Extroversion and Conscientiousness score can predicted from content-agnostic, and content-based descriptions of visual behaviour respectively. This illustrates that social networking site users can be psychologically profiled from collecting only 20 seconds of viewing behaviour. We discuss the impact to applications in user engagement during human-computer interactions and highlight potential privacy concerns.

## 5.2 Twenty Seconds to Know You? Classifying Personality from Visual Behaviour on Social Media

Tailoring content to appeal to the user's personality can promote consumer loyalty and engagement (Matz & Netzer, 2017). Similarly, appealing to an individual's personality can lead to increased conversion rates during online marketing campaigns, with personality-congruent personalised advertisements leading to up to 50% more purchases compared to non-personalised or personality-incongruent advertisements (Matz, Kosinski, Nave, & Stillwell, 2017). As such, the ability to quickly predict the personality of the user is of value to providers that wish to maximise the potential for users to engage with, and relate to, a wide range of services and content.

Online social networking sites (SNS) provide content that is socially and emotionally relevant to the user and enables users to connect, share content, and interact with others as part of a personally tailored experience. Machine learning techniques have been successfully applied to records of SNS behaviour to predict aspects of the user's private traits and attributes, such as their age, gender, political inclination, and personality (Kosinski, Stillwell, & Graepel, 2013). A recent meta-analysis identified that the self-reported 'Big Five' personality traits (McCrae & Costa, 2004) were the most commonly predicted individual characteristics from online digital traces, and that the Facebook platform was the most common SNS investigated (Settanni, Azucar, & Marengo, 2018). The authors found a moderate meta-correlation (0.34) between various digital traces and personality across 29 independent data sets, illustrating that an individual's personality is reflected in their online behaviour on Facebook (Settanni et al., 2018). However, currently existing methods of predicting a user's personality from SNS engagement require access to the user's detailed personal content and previous behaviour, often across months or years of use.

Due to the volume of data provided by eye tracking, a possible advantage of predicting a user's personality from their oculomotor behaviour is that accurate predictions may not require past knowledge of SNS behaviour, providing a stand-alone method to evaluate aspects of the user's personal attributes from a single interaction. Bargary and colleagues found, within a sample of 1,000 young adults, that an individual's eye movements during a variety of oculomotor tasks (e.g., following a moving object) provide a distinct and reliable 'oculomotor signature' that is stable across time (Bargary et al., 2017). This makes theoretical sense: visual behaviour, as a reflection of the spatial distribution of attention, is driven in part by our endogenous associations with features of the visual scene (Hayhoe & Ballard, 2005). We tend to look longer at visual stimuli which we find emotionally salient compared to those that we do not (Nummenmaa, Hyönä, & Calvo, 2006), and eye movements are influenced by individual attributes such as aspects of our personality (Rauthmann, Seubert, Sachse, & Furtner, 2012) and our cognitive biases (Armstrong & Olatunji, 2012). Furthermore, previous literature has identified that personality traits can be decoded from visual behaviour within both real-world locomotion (Hoppe, Loetscher, Morey, & Bulling, 2018) and screen-based (viewing a series of static images; Berkovsky et al., 2019) tasks. These examples suggest that our visual behaviour may provide a signal that reflects a range of our underlying individual traits and attributes.

Importantly, it is currently unknown whether an individual's private traits and attributes can be predicted from their visual behaviour upon their own SNS profile. This is a unique environment, in which the visual properties of the scene may vary dramatically across participants in terms of context (e.g., relevance) and content (e.g., video, text or images presented). To address this question, we allowed participants to browse their own Facebook News Feed (Centre, 2019) section while tracking their eye movements, and employ a machine learning approach to predict each of the Big Five personality traits (Openness, Conscientiousness, Extroversion, Agreeableness and Neuroticism McCrae & Costa, 2004). As there are many ways to characterise visual behaviour, we form five sets of features that each describe a particular aspect of oculomotor behaviour (outlined in Section 5.3.6) and independently assess the relative insight provided by each group into each of the Big Five personal traits. These feature groups form two broad categories that either describe overall eye movement statistics (e.g., as in; Hoppe et al., 2018), or, in a novel approach, the distribution of spatial attention across multiple types of content. As such, in a data-driven approach, we assess whether visual behaviour in relation to broad semantic categories (e.g., video, images, and text) is more informative of trait personality than statistical descriptions of oculomotor events when individuals browse their own SNS content.

## 5.3 Methods

Two hundred and nine participants were recruited from an undergraduate university cohort ($M_{age} = 20.45$, $SD_{age} = 3.38$, 169 Females) with age ranging from 18-51 years. All participants had normal or corrected to normal vision and owned a Facebook account. Demographic descriptors were not used as variables in this study. Participants took part in exchange for course credits or a £5 reward. Twenty-nine participants (25 females) were excluded due to software failure or having less than 80% valid eye position samples. Of the remaining 180 participants, 161 reported using the Facebook platform daily or weekly, with 14 reporting monthly usage and five yearly. Data collection was halted when over two hundred participants were collected, as we judged this an appropriate trade off between providing enough observations for the algorithm to learn from whilst considering the costs (time and monetary) associated with collecting laboratory data from a large cohort. The Institutional Research Ethics Committee granted ethical approval.

### 5.3.1 Eye Tracking

Each participant viewed their personal Facebook News Feed using the Internet Explorer browser (full-screen) upon a 23-inch TFT monitor (1920 x 1080). The viewing distance was 60cm, and eye movements were collected using a Tobii TX300 infrared eye tracker sampling at 300Hz, allowing for free movement of the head. Using the Tobii Studio software, a five-point calibration procedure was followed, with the experimenter conducting a visual inspection of the calibration accuracy before the task started. Stimuli were presented and data was collected within Tobii Studio. The Tobii Studio default I-VT filter (Window length: 20ms, Velocity threshold: $30^0$/s) was used to detect fixations (when the eyes are held steady upon the scene) and saccades (a fast movement to a new location in the visual scene).

### 5.3.2 Visual Stimuli

Each participant viewed (scrolled through) their own Facebook News Feed page for one minute, resulting in a unique visual scene for each individual. We asked participants to view the content as they would usually, with two caveats: to avoid messaging other users and avoid navigating away from the News Feed section of their Facebook account. The News Feed section hosts a constantly updating list of content displayed within a dynamically generated, scrolling central column. Example items include friend's status updates and shared content, along with advertisements and posts from public groups (for full details, see (Centre, 2019)). The Facebook platform provides several constraints upon the location of visual stimuli that are common across participants. In particular, the dynamically generated content is always displayed within bounding boxes of varying size (depending upon the content; e.g., multiple images, video, or text). Each bounding box is annotated with other user's reactions (likes, comments, etc.) directly below the content, and provides the user with space to engage in reacting to the content themselves (not used within this experiment). As such, this elicits a typical viewing pattern of observing a piece of content, followed by viewing the other user's reactions to that content.

### 5.3.3 Labelling Strategy

Using in-built Tobii Studio functionality, we obtain a record of the full web page content viewed, on to which visual behaviour has been mapped. Each web page is the product of concatenating multiple screens' worth of content that has been explored by the individual while scrolling through their own News Feed.

Upon manual inspection of the gaze data (comparing a video replay visualisation to the gaze data after being mapped onto a static image), it became apparent that the software would occasionally incorrectly stitch together the visual content as the participant scrolled. For example, it would correctly stitch together the first two pages worth of content but mis-align the third or fourth. By inspecting all participant data we found that, because of the above issue, the software did not always reliably capture the full one minute viewing duration, but always captured over twenty seconds correctly.

To ensure all participants were directly comparable, each participant's content was cropped to represent only the first twenty seconds of viewing behaviour. To protect the viewers' and their friends' anonymity, we obscured all content by blurring and, from earlier piloting, identify seven key types of content ('Create Post', 'Text Content', 'Image Content', 'Video Content', 'Hybrid Content', 'Interaction Elements', 'Comments') that frequently occur across participants (Figure 5.1). These aspects are manually labelled using the free software LabelImg (Tzutalin, 2015). Here, 'hybrid' refers to content whereby an image is overlaid with text.

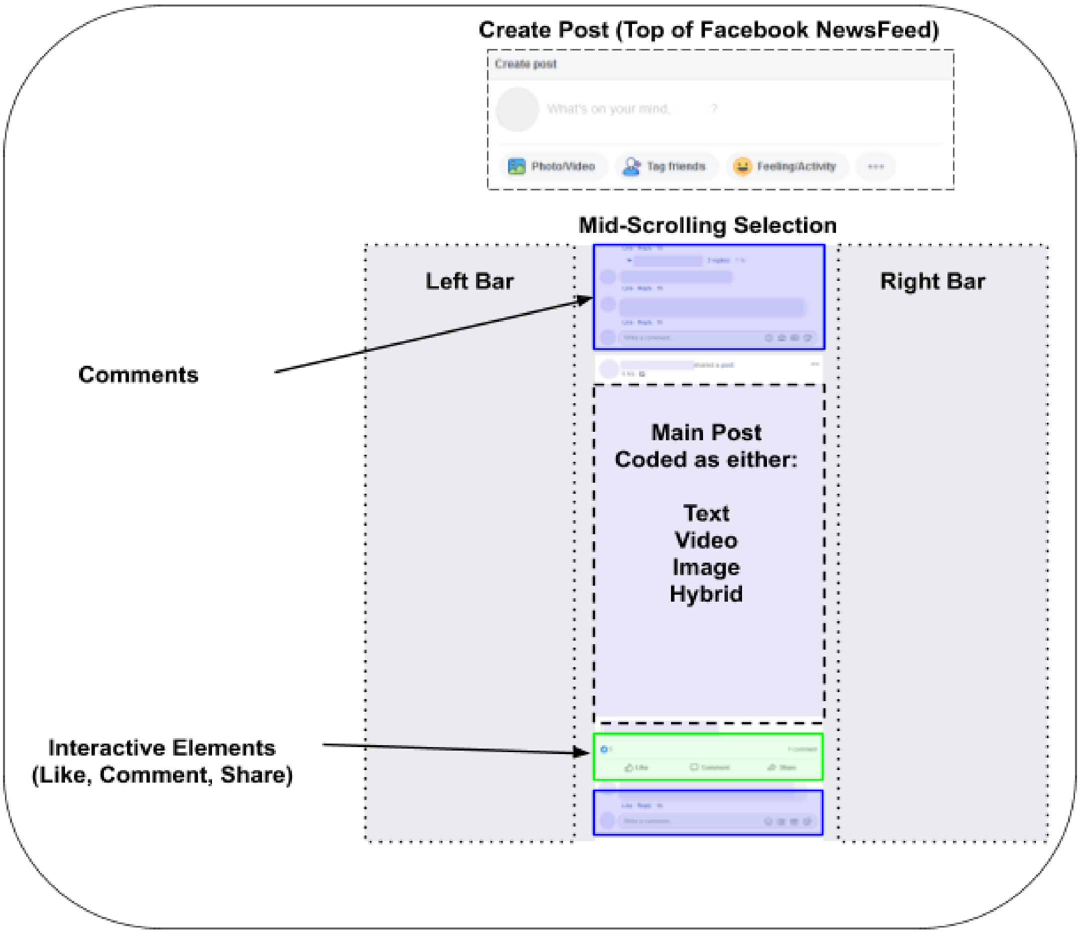


FIGURE 5.1: Labelling strategy for Facebook News Feed content. Only central elements within the mid-scrolling section were analysed. Coloured overlays are illustrative.

### 5.3.4 Questionnaire Materials

We collected demographic information from participants, including their age, sex, and personality traits using the NEO-FFI 60-item inventory (McCrae & Costa, 2004).

For each personality trait, a score between zero (minimum association with the trait) and 48 (maximum association with the trait) is calculated.

### 5.3.5 Machine Learning

We approached the prediction of personality traits as a supervised classification problem by splitting each trait into three categories (low, medium, and high) using an quantile-based binning strategy. Our binning did not result in exactly equal allocations due to discrete scores, thus, as a baseline, we report the highest accuracy and $F1_{macro}$ score possible by predicting all examples to be the majority class. We apply k-nearest neighbours (Guo, Wang, Bell, Bi, & Greer, 2003), ridge classification, (Friedman, Hastie, & Tibshirani, 2010) Support Vector Machines (Hearst, Dumais, Osuna, Platt, & Scholkopf, 1998), and naive Bayes classifiers (Zhang, 2004). We standardize each feature, ensuring each feature is upon the same scale, each having zero mean and a standard deviation of one. We utilise a nested (outer: five-fold, inner: five-fold) cross-validation procedure, selecting hyper parameters upon the $F1_{macro}$ score metric within the inner loops and collecting accuracy and $F1_{macro}$ scores across the outer loops. For models that have accuracy scores greater than baseline and an $F1_{macro}$ score above .33, we calculate the probability of finding the $F1_{macro}$ result by chance using label permutation (Ojala & Garriga, 2010), and correct these p-values for multiple comparisons using the Benjamini Hochberg (BH) procedure (Benjamini & Hochberg, 1995). Analysis is conducted using the Scikit-Learn library (Pedregosa et al., 2011).

### 5.3.6 Feature Engineering

In a novel approach, this work will be the first to assess the relative capability of using visual metrics that reflect spatial attention across broad categories of stimuli to accurately predict levels of the Big Five personality traits while participants view their own SNS content. Specifically, the algorithms are assessing the relative performance of visual metrics that reflect spatial attention, against more general statistical descriptions of eye movements. To understand which of these measures best predict each personality trait, we create separate groups of metrics ('feature groups' -

see Table 5.1) and evaluate the performance of the models trained upon them. Each feature group is based upon the first twenty seconds of viewing behaviour. Our first four feature groups explore area-of-interest (AOI) based metrics, whereby visual behaviour is grouped in correspondence to the category of content viewed upon the web page. Our fifth group is composed of statistical descriptions of eye movements that are not linked to any particular content. As a final sixth group, we introduce a control feature group describing the proportion of the page occupied, and frequency of occurrence, for each type of content labelled on the Facebook page. This allows us to understand if our oculomotor data is offering insight into each personality trait above and beyond that available from the characteristics of the user's Facebook News feed (according to the labelling strategy, as outlined in Section 5.3.3).

**AOI-Based Visual Metrics**

For the AOI-based metrics, we calculate the total fixation duration (TFD), number of fixations, and time to first fixation (TTFF) values in accordance with each of the content categories ('Create Post', 'Image', 'Text', 'Video', 'Hybrid', 'Comments', 'Interaction Elements'). For each participant, the visual scene is different, leading to a varying proportion of the page accounted for by each content category across participants. Furthermore, not all categories may be present for a particular participant. We take several steps to address these issues.

Firstly, when a content category never occurs for a particular participant (e.g., no videos appeared on their News Feed section within the first twenty seconds), they receive zero for TFD and number of fixations, and the maximum duration of 20 seconds for TTFF. To distinguish between instances when participants do not have the opportunity to encounter a type of content (and receive zero fixation duration/number) from when they do not fixate upon a content that is present, we supplement our AOI-based feature group by including the frequency of occurrence for each AOI type (denoted by 'with Frequency' in Table 5.1). However, as the stimuli size varies, this does not directly account for the amount (proportion) of the page occupied by each category. To address this, we reweighted the TFD and the number of fixation metrics by the proportion of the page each category (e.g., image, video) occupies (denoted by 'Proportional' in Table 5.1).

#### Non-AOI Based Visual Metrics

Inspired by previous literature (e.g., Baranes, Oudeyer, & Gottlieb, 2015; Hoppe et al., 2018), we create a separate set of metrics that represent overall statistical descriptions of fixation and saccadic behaviour across the Facebook News Feed. We name this group of features 'Eye Movement Statistics'. We consider fixations to have one attribute (duration) and saccades to have two attributes (duration, amplitude). For each attribute we calculate the sum, count, minimum, maximum, mean, standard deviation, range and interquartile range. We also include the mean progress in vertical screen-based coordinates per second (pixels per second) across the viewing duration, as an index of how quickly the participant progresses through their exploration of the web page. After creating these metrics, we removed minimum saccade duration as it was uniform across participants (defined by the fixation filter), and thus offered no discriminative power. Similarly, we removed the number of saccadic amplitudes as it is perfectly correlated with the number of fixations (amplitude is calculated between two fixations). Following these subtractions, a total of 23 features remained in the statistical feature group.

Finally, to understand the insight provided by knowing the page content alone (and not the visual behaviour), we included a control feature group consisting of the proportion of the page occupied, and the frequency of occurrence, for each content category (14 features).

TABLE 5.1: Feature Groups

| Feature Group | Number of Features |
|---|---|
| AOI | 21 |
| AOI Proportional | 21 |
| AOI with Frequency | 28 |
| AOI Proportional with Frequency | 28 |
| Eye Movement Statistics (EMS) | 23 |
| Page Content Info | 14 |

AOI: Area of Interest Based.

## 5.4 Results

### 5.4.1 Personality Questionnaire Distribution

Our personality questionnaire data contained one participant with one missing questionnaire item response, which was imputed as the mean of the remaining trait-congruent questionnaire items. The Shapiro-Wilk test for normality identifies that trait Conscientiousness scores show evidence of being non-normally distributed (W=.974, p=.002), and as such these scores may not be representative of the general population. No further traits demonstrated evidence of being non-normally distributed. Descriptive statistics for all traits, after splitting into low, medium, and high categories, are presented in Table 5.2.

TABLE 5.2: Descriptive Statistics for Big Five Personality Trait Scores (out of 48) split by category.

| Label | Low | Medium | High |
|---|---|---|---|
| | Mean (SD) | Mean (SD) | Mean (SD) |
| Openness | 23.95 (2.87) | 30.37 (1.48) | 36.74 (3.44) |
| Conscientiousness | 20.28 (4.73) | 29.94 (2.18) | 37.38 (2.83) |
| Extroversion | 21.97 (3.18) | 29.25 (1.51) | 35.05 (2.92) |
| Agreeableness | 24.41 (3.45) | 30.87 (1.25) | 37.37 (2.90) |
| Neuroticism | 17.12 (3.98) | 26.37 (1.69) | 34.63 (3.28) |

### 5.4.2 Social Media Content and Visual Behaviour

We note that the most frequently occurring type of content upon a user's own Facebook News Feed are interaction elements ('like', 'share', etc). Since these accompany each post, they also let us know that each participant viewed roughly 2-4 posts within their twenty second viewing duration. We report the average total fixation duration and the number of fixations metrics for each category (among participants where the content was shown) in Table 5.3.

TABLE 5.3: Area of Interest (AOI) Category Frequency and Fixation Behaviour

| Category | AOI Frequency | Duration* | Number** |
|---|---|---|---|
| | Mean (SD) | Mean (SD) | Mean (SD) |
| Comments | 0.88 (0.95) | 844 (822) | 3.55 (3.89) |
| Hybrid | 0.59 (0.87) | 2210 (1586) | 10.01 (7.17) |
| Image | 1.83 (1.33) | 2316 (1517) | 10.12 (6.58) |
| Text | 2.08 (1.70) | 2907 (1919) | 13.32 (8.81) |
| Video | 0.69 (0.87) | 1584 (1496) | 6.65 (6.19) |
| Create Post | 1 (0) | 258 (392) | 1.07 (1.31) |
| Interaction | 3.13 (1.59) | 642 (610) | 2.74 (2.29) |

*Average total fixation duration in milliseconds

**Average number of fixations

### 5.4.3 Classification Results

All significance values reported in this section are adjusted for multiple comparisons using the Benjamini Hochberg procedure (Benjamini & Hochberg, 1995). The best performance achieved across the feature groups for each personality trait is summarised in Table 5.4.

#### Eye Movement Statistics

For the Eye Movement Statistics feature set, we identify that only Extroversion can be predicted significantly better than chance, and this is achieved across multiple algorithms. The best performance comes from using a ridge classifier, followed by results from a linear support vector machine classifier.

#### AOI and AOI Proportional

AOI and AOI proportional feature sets were explored for each of the five personality traits. Only for Conscientiousness did we find that we were able to classify individual's trait category significantly better than chance. Specifically, for the AOI feature set our classifiers were modestly, but significantly, better at predicting Conscientiousness than chance using a ridge classification algorithm. For the AOI proportional feature set, Conscientiousness was also predicted significantly better than chance using a Ridge classification algorithm. The AOI feature sets (AOI and AOI Proportional) with frequency metrics performed modestly worse, with no classifiers significantly better than chance. We note that these results represent only a modest improvement over knowing only the page content information alone.

### Page Content Information

For the control feature group of page content information, we found that no traits were able to be predicted significantly above chance.

### Summary

Across the feature groups the traits of Conscientiousness and Extroversion were able to be predicted significantly better than chance. No feature group gave significant insight into the personality trait of Openness, Agreeableness or Neuroticism. We further explore the significant results presented within this section by trait category in section 5.4.4.

TABLE 5.4: Classifier Performance by Personality Trait

| Trait | Feature Group (Algorithm) | $F1_{Macro}$ (SD) | Accuracy (SD) | Baseline $F1_{Macro}$ (Accuracy) |
|---|---|---|---|---|
| Openness | EMS (SVM) | .327 (.020) | 38.9% (3.9) | 0.19 (40.0%) |
| Conscientiousness* | AOI Proportional (Ridge) | .398 (.079) | 42.8% (8.5) | 0.18 (36.1%) |
| Extroversion*** | EMS (Ridge) | .519 (.077) | 53.3% (7.3) | 0.19 (40.6%) |
| Agreeableness | Page Content Info (Ridge) | .340 (.058) | 38.9% (6.3) | 0.19 (38.9%) |
| Neuroticism | AOI (Naive Bayes) | .334 (.072) | 35.6% (6.7) | 0.17 (35.0%) |

* p<.05, ** p<.01, *** p<.001. AOI: Area of Interest. EMS: Eye Movement Statistics.

Ridge: One-vs-Rest Ridge Classification. KNN: K-nearest neighbors. SVM: Linear support vector machine

### 5.4.4 Exploratory Analysis

#### Classifier performance by trait category

For each significant model, we evaluate the $F1$ score for each trait category (low, medium, high) as shown in Table 5.5. To aid the reader, the standard deviation represents how stable the model's performance was across the five outer folds (e.g., how much performance varied with different training/test sets).

For Extroversion, the ridge classifier has a similar $F1$ score across each trait category, demonstrating a balanced classifier. It is relatively stable for the low and high categories, but variable for the medium category. In contrast, the support vector machine based classifier shows the best performance within the high category and worst within the low group, demonstrating an imbalance across the categories. Notably, the classifier is also the most variable in its performance for the medium category.

For Conscientiousness, the Ridge classifier based upon the AOI proportional features has similar $F1$ scores for both the medium and high categories, and substantially worse performance for the low category. The classifier's performance is most stable when predicting the high category, and most variable when predicting the medium category. For the Ridge classifier built upon the AOI feature set, the performance progressively improves with the quantile category ($low < medium < high$) and is also most variable for the medium category.

TABLE 5.5: Classifier Performance by Personality Category for Significant Models.

| Trait | Feature Group (Algorithm) | F1 Score (SD) Low | Medium | High | Accuracy |
|---|---|---|---|---|---|
| Extroversion | EMS (Ridge) | 0.51 (.069) | 0.50 (.168) | 0.54 (.095) | 53%*** |
| | EMS (SVM) | 0.37 (.092) | 0.40 (.153) | 0.48 (.063) | 43%* |
| Conscientiousness | AOI Proportional (Ridge) | 0.22 (.054) | 0.49 (.158) | 0.48 (.076) | 43%* |
| | AOI (Ridge) | 0.28 (.108) | 0.37 (.188) | 0.55 (.099) | 42%* |
| | Page Content Info (SVM)† | 0.34 (.101) | 0.40 (.091) | 0.43 (.059) | 40% |

* p<.05, ** p<.01, *** p<.001 corrected via Benjamini-Hochberg procedure. † Included for comparison.
Ridge: One-vs-Rest Ridge Classification. KNN: K-nearest neighbors. SVM: Linear support vector machine.
AOI: Area of Interest. EMS: Eye Movement Statistics.

To understand this further, for trait Conscientiousness and Extroversion we calculated how similar participants are within each category when responding to the twelve questionnaire items used to calculate the trait score. For Conscientiousness, the average euclidean pairwise distance between participants becomes smaller (participants respond more similarly) as the quantile-based category increases from low (4.04) to medium (3.9) and high (3.5). As such, we propose that individual's scoring low upon trait Conscientiousness represents a more diverse (less homogeneous) cohort than high scorers, which may result in a more challenging classification task. The low category also appears to be less homogeneous for trait Extroversion, with the highest average euclidean pairwise distance (4.29) compared to the medium (3.91) and high (4.12) Extroversion categories.

**Exploring the performance of classifiers trained upon features from multiple feature groups**

We have demonstrated the relative insight provided by each of the separate feature groups into each personality trait. To identify if combining features across groups (to make maximal use of the feature information available) can result in greater performance than any single feature group in isolation, we conduct a secondary exploratory analysis. We combine features from across the AOI Proportional, EMS

Statistics and Page Content Information feature groups and follow a similar nested-cross validation procedure. To reduce the dimension of this data, we first remove features that are over 95% correlated. Second, we employ recursive feature selection elimination within the inner cross-validation loop, seeking to find a minimum set of maximally informative features for each trait before fitting the classifier. Our results find equal or reduced performance across all traits, suggesting that combining the feature groups adds little trait-relevant information beyond that captured individually by the Eye Movement Statistics or AOI Proportional groups. As there are many alternative approaches to selecting an optimal subset of features, a full investigation is outside the scope of this paper. However, this is an important direction for future research.

## 5.5 Discussion

### 5.5.1 Overview

Online social networking sites (SNS) provide a rich and ecologically valid visual experience with a variety of content and information being presented. Previous literature has illustrated that various aspects of a user's online behaviour upon SNS, such as the distribution of 'likes' upon Facebook (Kosinski et al., 2013), or the content of text-based posts upon Twitter (Arnoux et al., 2017), can be used to predict aspects of an individual's personality. In a novel approach, we present evidence in support of our hypothesis that an individual's pattern of eye movements, whilst browsing their own Facebook News Feed section, is informative of aspects of their personality (Extroversion and Conscientiousness). Our research demonstrates that, with no prior knowledge, twenty seconds of visual behaviour upon an SNS site is enough to predict aspects of an individual's personality significantly better than chance.

### 5.5.2 Eye Movement Statistics

As the perceived task and type of content influences the expression of visual behaviour (Tatler, Wade, Kwan, Findlay, & Velichkovsky, 2010), we sought to understand how to best characterise visual behaviour in a way that reflects (is predictive of) personality. Within our paradigm, statistical descriptions of visual behaviour that are not tied to any particular content are more informative of trait Extroversion than descriptions of visual behaviour derived from responses to a particular category of stimuli (e.g., AOI-based metrics).Furthermore, we demonstrate that personality can be predicted from a substantially shorter time scale than employed within previous literature (e.g., 20 seconds vs Hoppe et al.'s 12.5 minutes and Berkovsky et al.'s 9.2 minutes). Together, this illustrates that statistical descriptions of oculomotor events are informative of trait Extroversion within extremely short time periods, even when each participant views a diverse range of visual stimuli. Our finding of Extroversion being linked to visual behaviour upon SNS content also expands upon the previous work of Rauthmann et al. (2012), who found that in abstract video-based stimuli Extroversion was linked to visual behaviour (being predicted by shorter dwelling times). Our finding could relate to the strong links between Extroversion and sociability (Plomin, 2010), which, given the nature of our social media stimuli, may have provided relevant visual content for evoking trait-congruent visual behaviour. That Extroversion appears to be rapidly manifested in statistical descriptions of oculomotor behaviour whilst viewing SNS content has substantial consequences for the application of personality-detection within applied settings, as this implies it is not essential to know the exact stimuli being presented.

### 5.5.3 Characterising Eye Movements in Relation to Types of Content

In a novel contribution to the literature, we identify that AOI-based metrics outperform statistical descriptions of visual behaviour when predicting trait Conscientiousness. Our results suggest that, when viewing SNS content, trait Conscientiousness is reflected in the way that the individual distributes their attention across different types of content within the visual scene. In considering why conscientiousness is the only trait presenting better results for the new AOI features (in comparison with the EMS results) we note that Conscientiousness is related to the ability for self-regulation during effortful control (Jensen-Campbell et al., 2002), with the individual's scoring higher upon Conscientiousness being more likely to maintain a consistent approach to the given task. In our paradigm, the task was to view their own Facebook NewsFeed, which provides the ongoing ability for the participant to decide to receive a new stimulus during the session (e.g., 'Do I look at the comments, or keep on scrolling?'). Thus, it may be that the participant's level of Conscientiousness influenced their style (e.g., systematic or more chaotic) of visually exploring the content, leading to Conscientiousness being reflected within the distribution of the participant's visual behaviour across the different content types.

We note that the performance of the classifiers predicting trait Conscientiousness varied in their performance across the low, medium, and high categories - with the lowest performance when predicting low scorers, and the most success in predicting high scorers. A possible explanation for the varying performance across the categories also comes from the nature of the trait itself, with Conscientiousness being positively associated with task completion, (Claessens, Eerde, Rutte, & Roe, 2010), and adherence (Molloy, O'Carroll, & Ferguson, 2014). Thus, we would expect Conscientiousness to influence both an individual's visual behaviour and their completion of the sixty-item personality questionnaire; with high scorers being more methodical (thus exhibiting similar behaviour) and low scorers being less principled (thus exhibiting more varied behaviour) in their approach. To explore this, we calculated how similar participants are within each Conscientiousness category when responding to the twelve questionnaire items used to calculate trait Conscientiousness. Our results support this interpretation, with the average Euclidean pairwise distance between participants becoming smaller (participants respond more similarly) as the quantile-based category increases from low to high. As such, we propose that an individual's scoring low upon trait Conscientiousness represents a more diverse (less homogeneous) cohort than high scorers, which is reflected within the labelling strategy and may result in a more challenging classification task.

### 5.5.4 Summary of Findings

Overall, we find that only the traits of Extroversion and Conscientiousness were able to be predicted significantly better than chance from our visual behaviour metrics. For the remaining personality traits, the descriptions of visual behaviour held little information above and beyond knowing the content upon the page. This is a different pattern of results than found within Chapter 4, where the traits of Extroversion, Agreeableness and Neuroticism were predicted significantly better than chance using similar metrics. It also appears to conflict with some previous findings such as

Berkovsky et al. (2019) who showed that characterising visual behaviour in response to multiple static images can be highly informative of trait personality ($>$ 61% accuracy upon the Big Five personality traits).

The difference in results may be attributable to a key methodological difference. In Berkovsky et al. (2019) and Chapter 4, visual behaviour was described according to each image seen, which, due to the images being identical across participants, were directly comparable. This allows the reasonable assumption that the observed variances in visual behaviour between participants are driven by individual differences, rather than the visual properties of the image (Itti & Koch, 2001). In contrast, in this study our AOI categories represent not a single identical image, but a diverse range of content, and items within a single category may vary in colour, spatial frequency, subject matter, and more. While this accurately reflects the complex variety of visual and social contexts present upon a fully featured SNS platform, the expression of visual behaviour is influenced by the properties of the visual stimuli (Hayhoe & Ballard, 2005; Itti & Koch, 2001; Tatler et al., 2010). As such, our design is likely to have introduced a substantial amount of variance in visual behaviour not directly linked to the user's personality, which increases the difficulty of the classification problem and may have led to reduced performance. This illustrates that it is perhaps not surprising that a different pattern of results were found, and raises questions regarding whether our results are directly comparable to studies utilising static free-viewing designs. We speculate that the findings of this study may suggest that models built upon descriptions of oculomotor behaviour in response to the static free viewing of images may not generalise well within applied SNS settings.

Finally, our choices for the AOI categories were informed by tasks identified as driving distinct visual behaviour (e.g., reading text, searching an image, or watching a dynamic scene; (Bulling, Ward, Gellersen, & Tröster, 2011; Tatler et al., 2010)), and aimed to capture visual behaviour in relationship to sufficiently broad categories as to be reasonably comparable across the majority of participants, while remaining sufficiently distinct to reflect a unique category of visual behaviour. As we kept our description of visual behaviour broad (regarding labelling of AOIs), the outlined technique could be applied to any web page and this is a direction for future research. However, we note that alternative category choices may lead to improved (or reduced) performance in classifying personality from visual behaviour. Future research may wish to explore which content categorisation schemes best capture trait-congruent visual behaviour and detail the circumstances under which content-bound descriptions of oculomotor behaviour become no longer informative of personality.

### 5.5.5 Practical Implications for Personality Prediction

To recap, past research has suggested that tailoring a product's advertising to appeal to an individual's personality can lead to increased conversion rates during online marketing campaigns (Matz et al., 2017), and promote consumer loyalty and engagement (Matz & Netzer, 2017). As such, it is desirable to be able to understand the personality of the user in order to maximise the potential for presenting them with engaging human computer interactions. However, current methodologies for evaluating personality either require extensive previous knowledge about the user's past interactions, (Matz & Netzer, 2017; Settanni et al., 2018), or are disruptive to a user's enjoyment of the experience (e.g., a user may not wish to conduct a questionnaire before engaging in an interaction). The eye tracking based techniques discussed within this paper provide a novel and non-intrusive method of predicting an individual's Extroversion and Conscientiousness category (low/medium/high) from a single twenty second interaction. This may support the development of socially aware human-computer interfaces, as users' personalities and visual behaviours are both associated with distinct information-seeking characteristics (Al-Samarraie, Eldenfria, & Dawoud, 2016).

This raises the question; if personality traits are predictable from eye movements whilst browsing social media, then what other personal information might also be decodable? Provided the personal attribute influences the way that the participant seeks and engages with the content upon the web page, we speculate that many of the attributes that are reflected in the digital footprints of individual's browsing social media sites may also be reflected in the participants visual behaviour. If true, this has wide ranging consequences to privacy as previous literature investigating which personal attributes can be decoded from digital footprints upon social media platforms has yielded a vast array of results, ranging from the ability to predict an individual's political inclination (Kosinski et al., 2013) to their gender and sexual orientation (Kosinski & Wang, 2017). Visual behaviour upon social media sites may also provide a uniquely direct insight into whether the individual is more prone to exploration (seeking novel content) or exploitation (capitalising upon known content). This may be predicted more directly from the participant's visual behaviour than would be possible from digital footprints alone, and avoids the need to predict an alternative proxy such as their Openness to Experience.

We note that with this emerging possibility comes new challenges for privacy. Targeted advertising is not always well received by the general public (Purcell, Brenner, & Rainie, 2012), and recent events such as Facebook knowingly violating both data privacy and anti-competition laws (Committee, 2019) only serve to highlight growing concerns over privacy. We suggest that future research investigates public perception towards the use of potentially covert methodologies aimed at predicting private traits and attributes. While the technology described here may not yet be ready for practical applications, the ubiquity of eye tracking devices is growing (Coldewey, 2020) and it is essential to develop rigorous guidelines for the ethical deployment of such technologies. This is especially relevant given that our research suggests it is not essential to know the exact stimuli being presented to the individual (e.g., as with our findings for Extroversion) when predicting their personality from visual behaviour. This reduces the demand for rigorous labelling and processing of the users' social media content, and may provide a privacy-preserving method of implicitly assessing an individual's personality.

Interestingly, while we found we were able to predict trait Extroversion and Conscientiousness, we were unable to classify participants above chance for trait Openness, Agreeableness, or Neuroticism within our paradigm. There, therefore, appears to be a performance trade-off with measuring visual behaviour over diverse stimuli upon such short time scales when compared to results from previous literature (Berkovsky et al., 2019; Hoppe et al., 2018).

However, whilst the time scale is a potentially limiting factor, there are other psychological interpretations as to why some personality traits were predicted above chance whilst others were not. A recent paper by Al-Samarraie et al. (2016) investigated whether specific personality traits influence how individuals seek out and process information in information seeking tasks (i.e., whilst using an online search engine). The authors investigated factual, interpretative and exploratory information seeking paradigms and found in all three that Extroversion, Agreeableness and Conscientiousness correlated with the number and total duration of fixations expressed by the individual. In contrast, Openness and Neuroticism were not correlated with any of the measured eye movements. Therefore, if we conceptualise browsing Facebook as a information search task, it is perhaps not surprising that our results indicate that Extroversion and Conscientiousness were able to be predicted significantly better than chance, whilst Openness and Neuroticism were not. This leaves the contradictory finding for Agreeableness, which was not predicted significantly better than chance within our study, yet was found to significantly correlate with eye movements in information search tasks (Al-Samarraie et al., 2016). Agreeableness is likely to influence the individual's behaviour when choosing whether to accept a particular source of information during a search task, which effectively biases the decision of when to accept that the search goal has been fulfilled and the task has been completed. However, whilst browsing Facebook in this study the participants were engaged in a freeviewing task and not searching for a set goal (i.e., piece of information) and there was no explicit objective to meet. As this is not a directed search, there was no need for participants to choose when to stop and accept the information as sufficient to fulfil the objective, which may be why the trait of Agreeableness was found by Al-Samarraie et al. (2016) in their paradigm, but not replicated within this study. Overall, our study's results suggest that browsing the Facebook Newsfeed is similar to information search tasks in reflecting trait Extroversion and Conscientiousness, but our design lacked the acceptance criterion that we speculate may be needed for the eye movements to be influenced by the individual's Agreeableness. This provides a key direction for future research, as experimentally manipulating the browsing task would allow the researcher to empirically investigate if the inclusion of an acceptance criterion is essential for trait Agreeableness to be accurately predicted from visual behaviour.

### 5.5.6 Conclusion

This study explored the ability of visual behaviour upon an SNS site to give insight into an individual's personality, in a situation where the classifier has no previous knowledge regarding the user's past behaviour upon the SNS platform. We demonstrate that within a single twenty second encounter aspects of the users personality can be predicted significantly better than chance. This highlights the possibility of a future where, with additional development, a provider may be able to tailor the presentation of its services or products to the user's attributes within a very short time frame. However, as the current performance of these classifiers is low, there may be situations in which visual behaviour metrics can be combined with existing data sources to increase performance when predicting personality traits. For example, previous literature has illustrated that existing records of an individual's behaviour upon SNS sites (e.g., likes and language use; Kosinski et al., 2013; Park et al., 2015) can be informative of personality. Future research may wish to explore alternative labelling strategies and the possibility of leveraging existing recordings of user interaction to compliment the methodologies outlined within this paper. This may lead to the increased performance required for practical applications.

## 5.6 Acknowledgements

We thank Jenna Morshead, Tatiana Tsaryuk, and Vrishni Kanhye for invaluable help in collecting and labelling the data set.

## References

Al-Samarraie, H., Eldenfria, A., & Dawoud, H. (2016). The impact of personality traits on users' information-seeking behavior. *Inf. Process. Manag. 53*(1), –. doi:http://dx.doi.org/10.1016/j.ipm.2016.08.004

Armstrong, T., & Olatunji, B. O. (2012). Eye tracking of attention in the affective disorders: A meta-analytic review and synthesis. *Clin. Psychol. Rev. 32*(8), 704–723. doi:10.1016/j.cpr.2012.09.004

Arnoux, P.-H., Xu, A., Boyette, N., Mahmud, J., Akkiraju, R., & Sinha, V. (2017). 25 Tweets to Know You: A New Model to Predict Personality with Social Media. In *AAAI Conference on Web and Social Media* (Vol. 11, pp. 472–476). Montreal, Canada: AAAI Press.

Baranes, A., Oudeyer, P. Y., & Gottlieb, J. (2015). Eye movements reveal epistemic curiosity in human observers. *Vision Res. 117*, 81–90. doi:10.1016/j.visres.2015.10.009

Bargary, G., Bosten, J. M., Goodbourn, P. T., Lawrance-Owen, A. J., Hogg, R. E., & Mollon, J. (2017). Individual differences in human eye movements: An oculomotor signature? *Vision Res. 141*, 157–169. doi:10.1016/j.visres.2017.03.001

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, *57*(1), 289–300.

Berkovsky, S., Taib, R., Koprinska, I., Wang, E., Zeng, Y., Li, J., & Kleitman, S. (2019). Detecting Personality Traits Using Eye-Tracking Data. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (pp. 1–12). CHI '19. doi:10.1145/3290605.3300451

Bulling, A., Ward, J. A., Gellersen, H., & Tröster, G. (2011). Eye movement analysis for activity recognition using electrooculography. *IEEE Trans. Pattern Anal. Mach. Intell. 33*(4), 741–753. doi:10.1109/TPAMI.2010.86

Centre, F. H. (2019). How News Feed Works. https://www.facebook.com/help/1155510281178725.

Claessens, B. J. C., Eerde, W. V., Rutte, C. G., & Roe, R. A. (2010). Things to Do Today . . . : A Daily Diary Study on Task Completion at Work. *Applied Psychology*, *59*(2), 273–295. doi:10.1111/j.1464-0597.2009.00390.x

Coldewey, D. (2020). Facebook, YouTube, Netflix and more get eye-tracking apps from Tobii.

Committee, C. S. (2019). *Disinformation and 'fake news': Final Report*. The Digital, Culture, Media and Sport Committee.

Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of statistical software*, *33*(1), 1–22.

Guo, G., Wang, H., Bell, D., Bi, Y., & Greer, K. (2003). KNN Model-Based Approach in Classification. In *On the Move to Meaningful Internet Systems* (pp. 986–996). doi:10.1007/978-3-540-39964-3_62

Hayhoe, M., & Ballard, D. (2005). Eye movements in natural behavior. *Trends Cogn. Sci. 9*(4), 188–194. doi:10.1016/j.tics.2005.02.009

Hearst, M. A., Dumais, S. T., Osuna, E., Platt, J., & Scholkopf, B. (1998). Support vector machines. *IEEE Intelligent Systems and their Applications*, *13*(4), 18–28. doi:10.1109/5254.708428

Hoppe, S., Loetscher, T., Morey, S. A., & Bulling, A. (2018). Eye Movements During Everyday Behavior Predict Personality Traits. *Front. Hum. Neurosci. 12*, 105. doi:10.3389/FNHUM.2018.00105

Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nat. Rev. Neurosci. 2*(3), 194–203. doi:10.1038/35058500

Jensen-Campbell, L. A., Rosselli, M., Workman, K. A., Santisi, M., Rios, J. D., & Bojan, D. (2002). Agreeableness, conscientiousness, and effortful control processes. *Journal of Research in Personality*, *36*(5), 476–489. doi:10.1016/S0092-6566(02)00004-1

Kosinski, M., & Wang, W. (2017). Deep Neural Networks Are More Accurate Than Humans at Detecting Sexual Orientation From Facial Images | Stanford Graduate School of Business. *J. Personal. Soc. Psychol. (in Press. 114*(2), 246–257.

Kosinski, M., Stillwell, D., & Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. *Proc. Natl. Acad. Sci. 110*(15), 5802–5805. doi:10.1073/pnas.1218772110

Matz, S. C., Kosinski, M., Nave, G., & Stillwell, D. J. (2017). Psychological targeting as an effective approach to digital mass persuasion. *Proceedings of the National Academy of Sciences*, *114*(48), 12714–12719. doi:10.1073/pnas.1710966114. eprint: https://www.pnas.org/content/114/48/12714.full.pdf

Matz, S. C., & Netzer, O. (2017). Using Big Data as a window into consumers' psychology. *Current Opinion in Behavioral Sciences*, *18*, 7–12. doi:10.1016/j.cobeha.2017.05.009

McCrae, R., & Costa, P. (2004). A contemplated revision of the NEO Five-Factor Inventory. *Pers. Individ. Dif. 36*(3), 587–596. doi:10.1016/S0191-8869(03)00118-1

Molloy, G. J., O'Carroll, R. E., & Ferguson, E. (2014). Conscientiousness and Medication Adherence: A Meta-analysis. *Annals of Behavioral Medicine*, *47*(1), 92–101. doi:10.1007/s12160-013-9524-4

Nummenmaa, L., Hyönä, J., & Calvo, M. G. (2006). Eye movement assessment of selective attentional capture by emotional pictures. *Emotion*, *6*(2), 257–268. doi:10.1037/1528-3542.6.2.257

Ojala, M., & Garriga, G. C. (2010). Permutation Tests for Studying Classifier Performance. *J. Mach. Learn. Res. 11*(Jun), 1833–1863.

Park, G., Andrew Schwartz, H., Eichstaedt, J. C., Kern, M. L., Kosinski, M., Stillwell, D. J., … Seligman, M. E. (2015). Automatic personality assessment through social media language. *J. Pers. Soc. Psychol. 108*(6), 934–952. doi:10.1037/pspp0000020

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., … Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res. 12*(Oct), 2825–2830.

Plomin, R. (2010). Extraversion: Sociability and Impulsivity? *Journal of Personality Assessment*, *40*(1), 24–30. doi:10.1207/s15327752jpa4001_6

Purcell, K., Brenner, J., & Rainie, L. (2012). *Search Engine Use 2012*. Pew Research Center.

Rauthmann, J. F., Seubert, C. T., Sachse, P., & Furtner, M. R. (2012). Eyes as windows to the soul: Gazing behavior is related to personality. *J. Res. Pers. 46*(2), 147–156. doi:10.1016/j.jrp.2011.12.010

Settanni, M., Azucar, D., & Marengo, D. (2018). Predicting Individual Characteristics from Digital Traces on Social Media: A Meta-Analysis. *Cyberpsychology, Behav. Soc. Netw. 21*(4), 217–228. doi:10.1089/cyber.2017.0384

Tatler, B. W., Wade, N. J., Kwan, H., Findlay, J. M., & Velichkovsky, B. M. (2010). Yarbus, eye movements, and vision. *Iperception. 1*(1), 7–27. doi:10.1068/i0382

Tzutalin. (2015). LabelImg.

Zhang, H. (2004). The Optimality of Naive Bayes. *AAI*, *1*(2), 6.

# Chapter 6

# The effect of labelling strategy upon predicting private attributes from visual behaviour

## 6.1 Abstract

When applying machine learning techniques to predict personality from digital records, the traits are often split arbitrarily into two or three categories using a quantile binning strategy. We propose an alternative method that provides insight into the number of categories to form, while making fewer assumptions about the underlying personality data. Using 180 undergraduates' Big Five personality scores, we illustrate that our approach leads to categories which are more distinct and internally cohesive than those created using a quantile based approach. The splitting strategy employed influences classifier performance for some traits more than others. In the final study, we apply our method when predicting personal attributes outside of personality (sex, political inclination, self-esteem, and narcissism). In a novel addition to the literature, we find that aspects of an individual's self-esteem, narcissism and sex can be predicted significantly better than chance. We note that future research may wish to explore the influence of the visual domain upon the ability to predict private attributes from visual behaviour and highlight relevant privacy concerns.

## 6.2 Introduction

A rapidly accumulating literature has focused upon how records of online user behaviour in response to social networking site (SNS) content can reflect aspects of the user's private traits and attributes. For example, the landmark study by Kosinski, Stillwell, and Graepel (2013) demonstrated that classifiers trained upon the distribution of a user's 'likes' upon the Facebook platform were able to predict the individual's political inclination, personality, and a wide range of other personal attributes significantly better than chance. Following this, the field of predicting personal attributes from records of online behaviour upon SNS has flourished. Examples include predicting personality from lexical attributes upon text-based platforms such as Twitter (Arnoux et al., 2017), or from the type of photographs shared upon Instagram (Ferwerda, Schedl, & Tkalcic, 2015) and Facebook (Eftekhar, Fullwood, & Morris, 2014; Segalin et al., 2017).

In this paper, we focus upon how eye movements can be used to gain insight into an individual's personality whilst browsing social media content. A linear association between visual behaviour and personality attributes had previously been found whilst participants view abstract animations (Rauthmann, Seubert, Sachse, & Furtner, 2012), with the association being thought to emerge from the individual's style of seeking visual information that is driven by their cognitive biases and preferences (Hayhoe & Ballard, 2005). More recent works have found that aspects of an individual's personality can be predicted from descriptions of their visual behaviour across a series of evocative images (Berkovsky et al., 2019) and in a more ecological setting of everyday locomotion (Hoppe, Loetscher, Morey, & Bulling, 2018) using machine learning techniques within a supervised classification setting. Finally, in a previous study, we demonstrated that twenty seconds of Facebook NewsFeed browsing is enough for aspects of an individual's personality to be predicted significantly better than chance from their visual behaviour (Chapter 5). The advantage of employing a machine learning approach when modelling the association between visual behaviour and personal attributes, is that it allows the researcher to capture complicated, non-linear associations between the input (descriptions of visual behaviour) and the desired personal attribute (e.g., personality). In the case of predicting a personality trait score, a common technique is to split the continuous trait into multiple categories, as this enables the researcher to apply learning algorithms that would otherwise be incompatible with the regression problem. This signifies a shift in priorities whereby the researcher accepts less explanatory power to gain more predictive power (Yarkoni & Westfall, 2017). A common technique for converting continuous trait scores into categorical outcomes is by applying a quantile-based split (e.g., to form low, medium, and high categories as in: Berkovsky et al., 2019; Hoppe et al., 2018). This technique seeks to allocate an equal number of observations to each category by dividing the underlying probability distribution into equal parts. This is justified to a certain extent, as learning from imbalanced data brings special considerations that can produce misleading results when using common performance metrics such as accuracy (Alpaydin, 2014). As such, ensuring an equal number of observations in each category simplifies the training and evaluation of the classification algorithms. However, the quantile split approach makes three strong assumptions; one is that the number of naturally forming categories present within the data is already known, the second being that these categories account for an equal amount of the probability distribution (e.g., that each outcome is equally likely), and finally that the underlying sample is representative of the population and thus the categories will generalise well to new samples of that population.

The assumptions outlined are important as personality traits are thought to follow a normal distribution within the general public (McCrae & Costa, 2004), and when using a quantile split approach in representative samples, this forces the middle category to span a much smaller range of values than the low or high categories. This is a unnatural way of encapsulating whether someone has a 'low', 'medium' or 'high' affinity to a particular trait. Furthermore, employing this approach to category forming within studies that employ a small sample size (e.g., N=21: Berkovsky et al., 2019), may lead to the evaluation of classifiers upon outcomes that are highly sample-specific, and may not generalise well to the broader population.

Given the challenges with quantile splits, in this paper we explore the possible benefits of employing a data-driven approach to both choosing the number of categories to create and assigning observations to these categories. We suggest two unsupervised learning techniques that may help the researcher form more appropriate categories, *k*-means and agglomerative clustering. Clustering is a form of unsupervised learning, whereby we attempt to find natural groupings of observations based upon their similarity across a set of defined dimensions (Alpaydin, 2014). In our case, we have a special instance of clustering where we have a single dimension (the trait score) which we wish to segment to form categories. This simplifies the clustering problem as unlike clustering within multiple dimensions, it means that the solution is relatively invariant to the distance metric employed (Alpaydin, 2014).

In *k*-means clustering, the algorithm attempts to find a given number of clusters (denoted as *k*) by minimising the amount of variance across observations within each cluster. However, the technique requires that the researcher specify the number of clusters, which is often not known in advance. Agglomerative, or hierarchical, clustering is an alternative approach that starts with clusters consisting of one observation, and successively merges the clusters based upon a similarity metric until all observations are joined to form one cluster (Alpaydin, 2014). The resultant structure of nested clusters can be visualised via a dendrogram and may provide insight into the optimal number of clusters to form. However, as the hierarchical algorithm is constrained by previous merge decisions (e.g., it can not separate observations that have previously been assigned to a cluster), *k*-means clustering can often find a solution that further minimises the intracluster distance for a given number of clusters. As such, it is advisable to use hierarchical algorithms to provide visualisations that help decide upon the optimal number of categories, and *k*-means to find the best version (e.g., which minimises the intracluster distance) of those categories (Alpaydin, 2014). Given this, we propose a two-step technique whereby the researcher first visualises the dendrogram using agglomerative clustering to gain an insight into the underlying structure of the data, and then uses this insight to select *k* when forming categories using the *k*-means algorithm.

The described technique provides two key benefits over quantile-binning strategies when predicting personal attributes (e.g., personality) from visual behaviour. Firstly, we make fewer assumptions about the underlying data, as the method does not make the assumption that each category is equally likely. Secondly, we guide the researcher by providing an insight into the number of clusters to form. We propose that embracing a data-driven approach to category forming, which makes fewer assumptions about the prevalence of the trait categories, may lead to categorical boundaries that better represent the underlying data. While the choice of how many categories to form is still subjective and based upon the researchers intuitions (e.g., interpretation of the dendrogram structure), we propose that with our method the evidence now is available to evaluate the researcher's choice. Furthermore, by ensuring that the categories map onto the natural structure of the data, assuming a true underlying association exists between the input and output, we would expect that researchers may also find that their classifier's performance increases. We test this hypothesis upon an existing sample of participants (N=180) in predicting each of the big five personality traits (McCrae & Costa, 2004) from visual behaviour upon their own Facebook NewsFeed (Chapter 5).

To summarise, the prediction of private attributes, such as personality, from oculomotor behaviour (e.g., Baranes, Oudeyer, & Gottlieb, 2015; Berkovsky et al., 2019; Hoppe et al., 2018; Rauthmann et al., 2012) often follows a common formula. First, the participant is asked to take part in a visual task while their eye movements are recorded. Second, the participant is asked to fill out a demographic form, within which is a series of questions designed to assess the individual's private traits and attributes. In more recent literature, the machine learning task has been framed as a supervised learning paradigm, whereby known labels (as assessed via the demographics questionnaire) are used to learn a function that maps the input (e.g., eye movement descriptors) to the trait or attribute the researcher desires to predict. Thus, in our first study, we wish to understand how well the quantile-based strategy commonly employed within the literature represents the underlying data (e.g., does it group similar observations into the same category while maximising the difference between categories) and whether our clustering algorithm approach can improve upon this. In our second study, we apply this technique upon an existing data set linking eye movements to personality (detailed in Chapter 5), to understand how the labelling strategy may influence classifier performance. We characterise the performance of classifiers predicting the quantile-based and cluster-based personality labels upon the metrics of accuracy and $F1_{macro}$ score. In our third study, in a practical application of the proposed methodology, we investigate whether the same SNS-based data set can give insight into additional traits and attributes (political inclination, narcissism and self-esteem) that are not part of the main hypothesis presented in the previous work. As part of this final study, we also evaluate whether visual behaviour can give insight into personal attributes that have known categorical outcomes. We investigate the attributes of sex ('male'/'female'), and whether the individual has voted in the last election ('no'/'yes').

## 6.3 Methods and Materials

### 6.3.1 Demographics

Using the same data as reported in Chapter 5, one hundred and eighty participants ($M_{age} = 20.48, SD_{age} = 3.57$, 144 females) were obtained from a university cohort. All participants had normal or corrected to normal vision and owned a Facebook account, and most ($N = 161$) used Facebook daily. The Institutional Research Ethics Committee granted ethical approval.

### 6.3.2 Visual Stimuli

Participants viewed their own Facebook Newsfeed for one minute in a free-viewing paradigm. This leads to each participant viewing a unique visual stimulus. The first twenty seconds of each participant's viewing duration is analysed in this paper, being captured as a screenshot that was manually annotated. We label each of the posts within the central column of the participant's Facebook NewsFeed as belonging to one of seven categories ('Create Post', 'Image', 'Text', 'Video', 'Hybrid', 'Interactive Elements', 'Comments') and draw corresponding areas of interest (AOI) around each item. Further details, such as the frequency of occurrence for each of the content categories, are available to the interested reader in the original paper (see Chapter 5).

TABLE 6.1: Summary of Feature Groups

| Feature Group | Number of Features |
|---|---|
| AOI | 21 |
| AOI Proportional | 21 |
| AOI with Frequency | 28 |
| AOI Proportional with Frequency | 28 |
| Eye Movement Statistics (EMS) | 23 |
| Page Content Info | 14 |

Reproduced with permission from Chapter 5

### 6.3.3 Feature Groups

A summary reproduced from Chapter 5 is presented in Table 6.1, which outlines the six sets of features used when predicting private attributes from visual behaviour in experiments two and three. Four of these feature groups (e.g., 'AOI', 'AOI Proportional') describe visual behaviour in relation to different types of content (e.g., 'Image', 'Text'). For all content categories and prevalence metrics, please refer to the original paper. We refer to these feature groups as being area of interest (AOI) based. One set of features (eye movement statistics, or 'EMS') refers to statistical descriptions of fixations (when the eye is still) and saccades (where the eye is moving to a new location) such as their mean duration, or amplitude. The sixth feature group documents the frequency of occurrence and the proportion of the page occupied by each type of content during the viewing duration. A detailed description of each feature group can be found in the original paper.

### 6.3.4 Clustering Procedure

We employ the $k$-means++ and agglomerative clustering algorithms as implemented within the sklearn package (Pedregosa et al., 2011). For the agglomerative clustering algorithm, we select the ward linkage function (Ward, 1963) which seeks to minimise the intracluster distance while weighting the distance equally in all directions. The silhouette score is calculated for each observation as $\frac{(b-a)}{max(a,b)}$ where $a$ is the mean intra-cluster distance and $b$ is the distance between the observation and the nearest observation in a cluster that it does not belong to (Ward, 1963). Silhouette score varies from -1 (worst) to +1 (best), and is calculated per observation. To receive a high score, the observation must be more similar to observations within its category than to observations outside of its category (for a more thorough treatment, see Rousseeuw, 1987). We standardize the trait scores before clustering and report the mean silhouette score across samples.

### 6.3.5 Personality Questionnaire

Participants completed the NEO-FFI sixty-item personality questionnaire (McCrae & Costa, 2004). Participants respond to each item upon a five-point Likert scale (0-4), ranging from 'Strongly Disagree' to 'Strongly Agree'. Twelve items are dedicated to each of the five traits (Openness, Conscientiousness, Extroversion, Agreeableness, Neuroticism), with six of each twelve items being reverse scored (e.g., such as that answering 'Strongly Disagree' is the most trait congruent response). Each participant's score for each trait ($min : 0, max : 48$) is calculated as the sum of the 12 trait-congruent questionnaire item scores.

## 6.4 Study One: Hierarchical Clustering

In this first study, we evaluate the categories formed when splitting each personality trait using the quantile-based technique employed within previous literature (Berkovsky et al., 2019; Hoppe et al., 2018), compared to our proposed technique. We consider that instances within each category should be more similar to each other than observations outside of the category, and thus evaluate the results based upon the metric of silhouette score (Rousseeuw, 1987). First, we directly compare the quantile split and $k$-means strategies by reporting the mean silhouette score for each when splitting each personality trait into two, three, or four categories. Secondly, we then illustrate how visualising the dendrogram gained from agglomerative clustering can provide insight into the underlying structure of the data and act as a visual aid when evaluating the selection of $k$. Finally, we combine the two techniques to form our two-step procedure, using the dendrogram to choose a number of categories that we judge to best reflect the underlying data ($k$), and then applying $k$-means clustering to find a good solution for the given number of categories. In this final evaluation, for each trait, we compare the solution gained from using our two-step procedure against the blind application of a three-class quantile split. For brevity, we report the full procedure for one example trait, with visualisations (e.g., dendrogram visualisations based upon the agglomerative clustering technique) for the remaining traits available within the supplementary materials (see supplementary materials B).

### 6.4.1 Results - Three Class Split

When creating between two and four categories, k-means consistently equals or outperforms the quantile based approach as seen in Figure 6.1. The quantile split strategy performs best when forming two categories, after which the strategy rapidly leads to smaller silhouette scores (indicating less appropriate categories). In contrast, the k-means approach maintains a similar silhouette score as the number of categories increases. As such, the magnitude of the difference between the strategies generally increases with the number of categories.

When focusing upon the common three-class split instance, we show the boundaries (e.g., range of each personality trait category) and support (number of observations in the category) for each technique in Figure 6.2. We observe that the *k*-means strategy generally creates a broader medium category than the quantile split method. Overall, in the three-case instances, employing a k-means clustering approach leads to a small overall improvement in the mean silhouette score across all personality traits ($Mean = .068, Std = .031$). Importantly, the magnitude of improvement varies across personality traits, with the most improvement seen for Openness (Graph 6.3). This illustrates that the k-means clustering technique consistently equals or outperforms the quantile split approach in creating categories that are internally cohesive and distinct from each other.

### 6.4.2 Results - Using Hierarchical Clustering to Inform the Class Split

As previously mentioned, the silhouette score provides a goodness-of-fit metric and is calculated sample-wise. The largest average silhouette width generally indicates the best number of clusters (Mirkin, 2011), which allows us to use the silhouette scores outlined in Figure 6.1 to interpret the best number of categories for each trait.

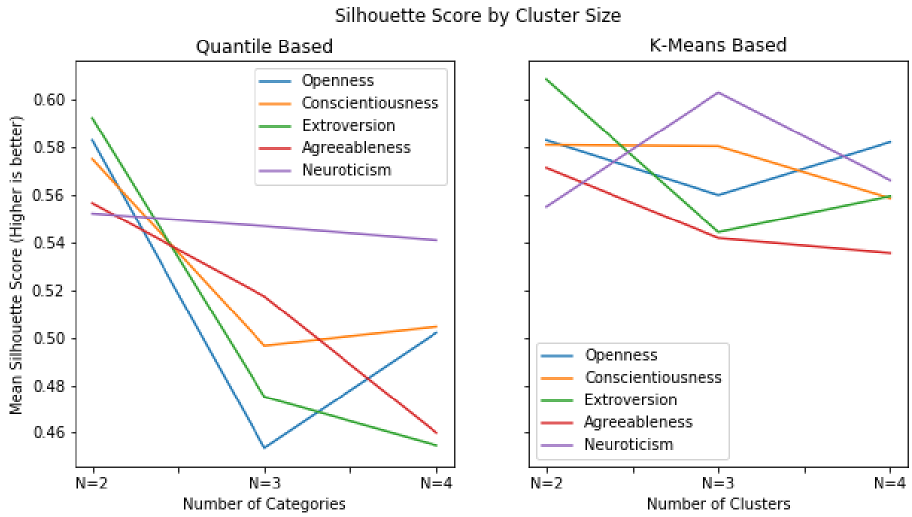


FIGURE 6.1: Silhouette Score by Cluster Size

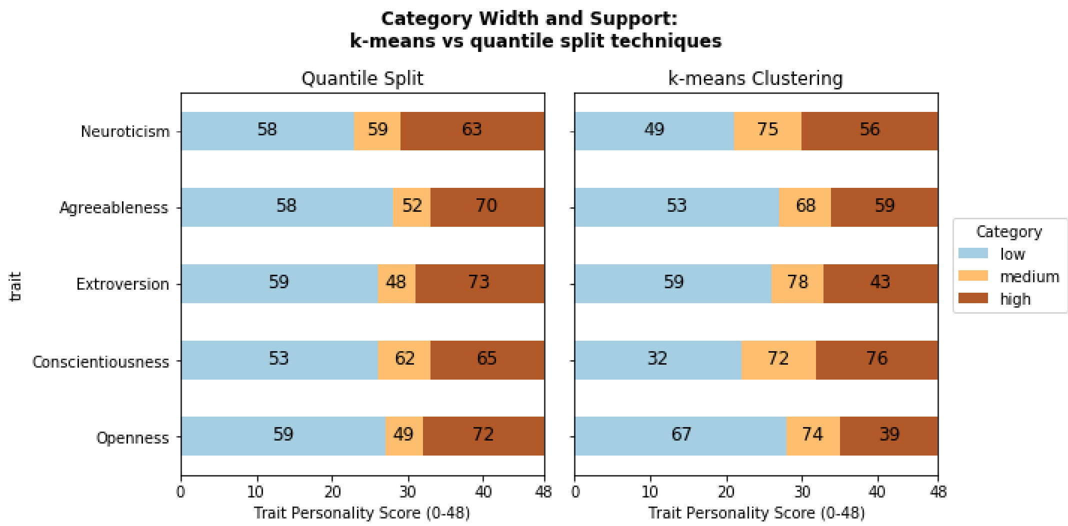


FIGURE 6.2: Category Width by Trait - Kmeans and Quantile Split.

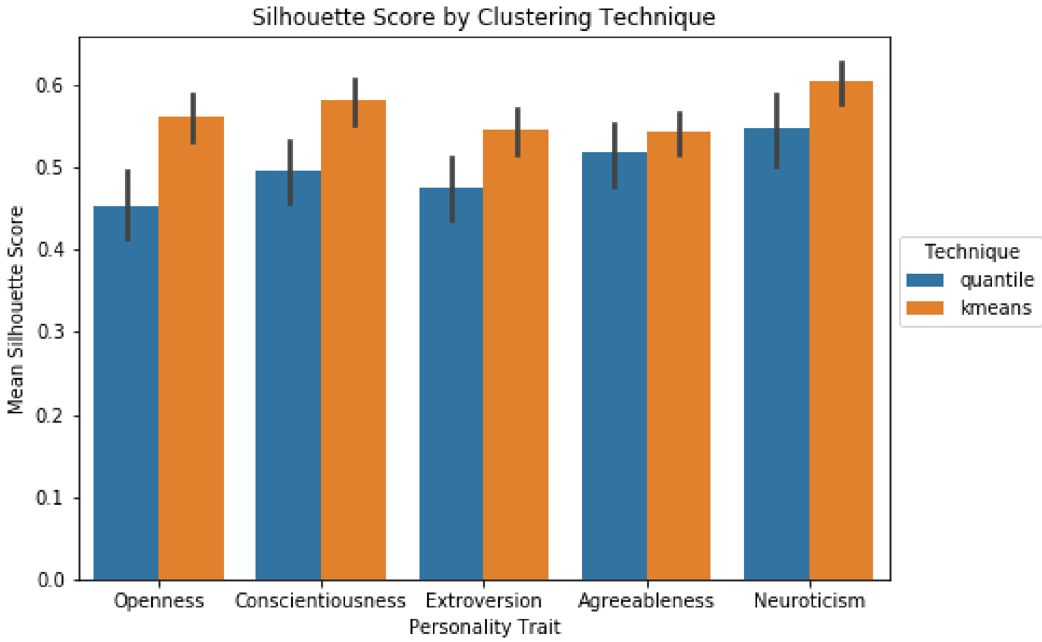


FIGURE 6.3: Silhouette score by clustering technique for the three-class instance. Error bars: 95% confidence interval.

However, rather than iterating through various values of $k$, we can further refine our methodology by using an alternative technique - agglomerative clustering. This provides a way of cross-examining likely candidates for the number of categories to form. First, we visualise the dendrogram formed by successively agglomerating samples in a bottom-up manner, as this represents the algorithm's interpretation of the inherent structure within the data. Our objective is to draw a horizontal line across the y-axis (representing the ward linkage distance) such as to form categories that respect the underlying data structure. For brevity, we demonstrate the entire procedure for Conscientiousness in detail, before comparing our approach to the three-category quantile split strategy.

The results of the agglomerative clustering for trait Conscientiousness are shown in Figure 6.4. As we have clustered upon the single continuous dimension of conscientiousness score, each cluster is theoretically valid (e.g., represents a region of the scale) and we are searching for a trade-off between representing the underlying data and avoiding the formation of many underrepresented categories. For example, in Figure 6.4, drawing a horizontal line at $y = 1$ would form many categories that each

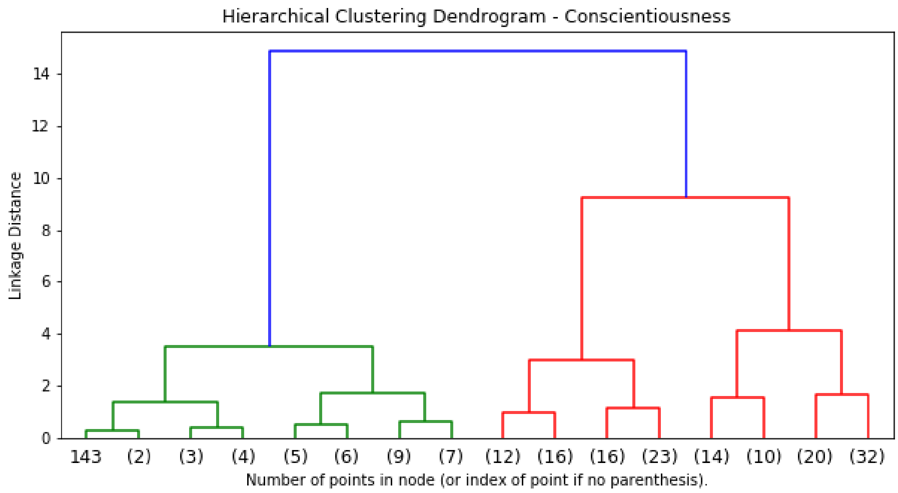


FIGURE 6.4: Hierarchical Clustering Dendrogram for trait Conscientiousness. Ward linkage function with the top three nodes shown.

span a small region of the scale and contain few instances. In this instance, we consider that $y = 6$ provides a reasonable trade-off whereby three categories are formed with the smallest accounting for 20% of the samples ($N_{support} = [76, 67, 37]$). This replicates the findings of the previous analysis whereby $N = 3$ yielded the highest average coefficient for Conscientiousness (Figure 6.1), and gives an additional insight into the structure within the data. As such, we select $k = 3$ and assign the observations to these categories using the $k$-means algorithm.

We repeat this process for each of the five personality traits to identify the optimal number of categories for each trait, with dendrograms for each being available in the supplementary materials (Supplementary materials B). At the end of this process, for each trait, each observation has been assigned to a category using k-means clustering. The number of categories formed for each trait has been selected by visual inspection of the dendrogram produced by agglomerative clustering.

As we wish to understand if following this procedure has led to a significant improvement in the quality of the categories formed, for each trait, we evaluate whether there has been a significant improvement in the average silhouette score compared to the blind application of a three-category quantile split (as is common

TABLE 6.2

| Trait | Suggested Category N | Silhouette Score - Mean (SD) Clustering Method | Quantile Method† |
|---|---|---|---|
| Openness*** | 2 | .583 (.162) | .454 (.284) |
| Conscientiousness*** | 3 | .580 (.192) | .497 (.261) |
| Extroversion*** | 2 | .609 (.161) | .475 (.260) |
| Agreeableness | 3 | .542 (.174) | .518 (.252) |
| Neuroticism* | 3 | .603 (.177) | .547 (.289) |

*p<.05, **p<.01, ***p<.001

† Quantile-based silhouette score for the three category case

in the literature). We find the process is worthwhile, as it led to a significant improvement in the average silhouette for each personality trait (compared to the three-category quantile split), except for Agreeableness (see Table 6.2).

### 6.4.3 Interpreting Clustering Results

We proposed that employing a *k*-means clustering approach, when carefully evaluated (e.g., cross-examined using agglomerative clustering), provides a data-driven method for forming categories when seeking to predict trait personality scores as a categorical outcome. This approach offers two main advantages over a quantile split method (e.g., as used in: Berkovsky et al., 2019; Hoppe et al., 2018). Firstly, we do not make the strong assumption that each category is equally likely, as it seeks to divide the range of scores so as to produce categories that are internally homogeneous and distinct from each other. This allows each category to have a unique probability that reflects the underlying data. Secondly, our approach provides guidance upon the optimum number of categories to form within each personality trait. Our hypothesis was that this would lead to the informed creation of personality score categories that better represent the data.

Our results support this suggestion, with the *k*-means approach consistently outperforming the quantile split technique within our sample. The only exception to this appears to be trait Agreeableness, where there the improvement does not reach significance. This demonstrates that occasionally the quantile-split approach results in equivalent outcomes (e.g., when the data matches the underlying assumption of equally probable categories), but often this is not the case and can be improved upon. We note that the proposed clustering technique still involves subjective decision making as there is no deterministic way to evaluate the optimal splitting strategy (Steinley & Brusco, 2011). However, the technique outlined here allows the reader to form their own intuitions as to the validity of the authors' approach, which provides a distinct advantage over the blind application of a three-category quantile split. In the next study, we evaluate whether this leads to improvements in classifier performance.

## 6.5 Study Two

The association between visual behaviour and personality is thought to arise due to the individual's personal preferences biasing the distribution of their visual attention. For example, we tend to look longer at stimuli which evoke strong positive or negative affective responses compared to neutral stimuli (Nummenmaa, Hyönä, & Calvo, 2006), and there is evidence to demonstrate that manipulating how long someone views particular content can influence their subjective evaluations (Shimojo, Simion, Shimojo, & Scheier, 2003). As such, it may be useful to conceptualise the distribution of visual attention as representing the individual's attempt to extract rewarding information from the scene, where reward is defined by both the relevance to the current task (Tatler, Wade, Kwan, Findlay, & Velichkovsky, 2010) and the user's personal preferences (Shimojo et al., 2003). In this study, we investigate whether the choice of labelling strategy influences classifier performance. To test this, we compare the performance of classifiers when predicting either the quantile-split based labels or the *k*-means based labels as outlined in study one. To allow direct comparison, we compare the classifier performance for both the two and three class instances (e.g., 2-class quantile split versus 2-class *k*-means, 3-class quantile split versus 3-class *k*-means). We propose that forming personality categories that accurately represent the underlying data should lead to improved performance for classifiers predicting this outcome. As our previous study found that both Conscientiousness and Extroversion could be predicted significantly better than chance from visual behaviour, and we predict that classifier performance for these traits will improve when using the *k*-means categories.

### 6.5.1 Methods

We evaluate the performance of the classifiers in predicting an individual's trait category from their visual behaviour while browsing their own social media site content. In study one, inspecting the dendrogram for Extroversion indicated that two categories better represented the underlying data, while three categories better represented Conscientiousness (see section 6.4). As such, we compare the performance of classifiers when predicting each personality trait category as defined either by the quantile split, or k-means clustering labelling strategy for both the two and three class instances. We note that the *k*-means technique does not make the assumption that each category is equally likely, and thus is not constrained to equally distributing instances across the categories. This leads towards an imbalanced learning paradigm where metrics such as accuracy may be misleading when evaluating classifier performance (Alpaydin, 2014). In the next section we discuss the steps taken to address this.

#### Machine Learning

We evaluate our classifiers upon the metrics of accuracy and $F1_{macro}$ score, using a nested cross-validation procedure (5-fold cross-validation for both inner and outer loops). We evaluate the feature groups (see section 6.1) using the same algorithms as outlined within Chapter 5, and report the best performance achievable when the categorical labels (e.g, low, medium, high) are assigned using the quantile split technique versus the k-means clustering technique. The $F1_{macro}$ metric is calculated as the unweighted mean of the $F1$ score for each category, where the $F1$ score is the harmonic mean of precision and recall for the given category (for more information, see: Alpaydin, 2014; Fawcett, 2006). For example, in our three-category case, $F1_{macro}$ would be calculated as $\frac{F1_{low}+F1_{medium}+F1_{high}}{3}$. As such, macro averaging ensures that the classifier must perform well across the categories to receive a high score. Accuracy is calculated as the sum of the correctly categorised instances divided by the number of instances. In the binary case, we report the standard $f1$ score. As $f1$ metrics are more robust to class imbalance, for the best classifier for each trait, we report the probability of finding the given $F1_{macro}$ score via permutation testing

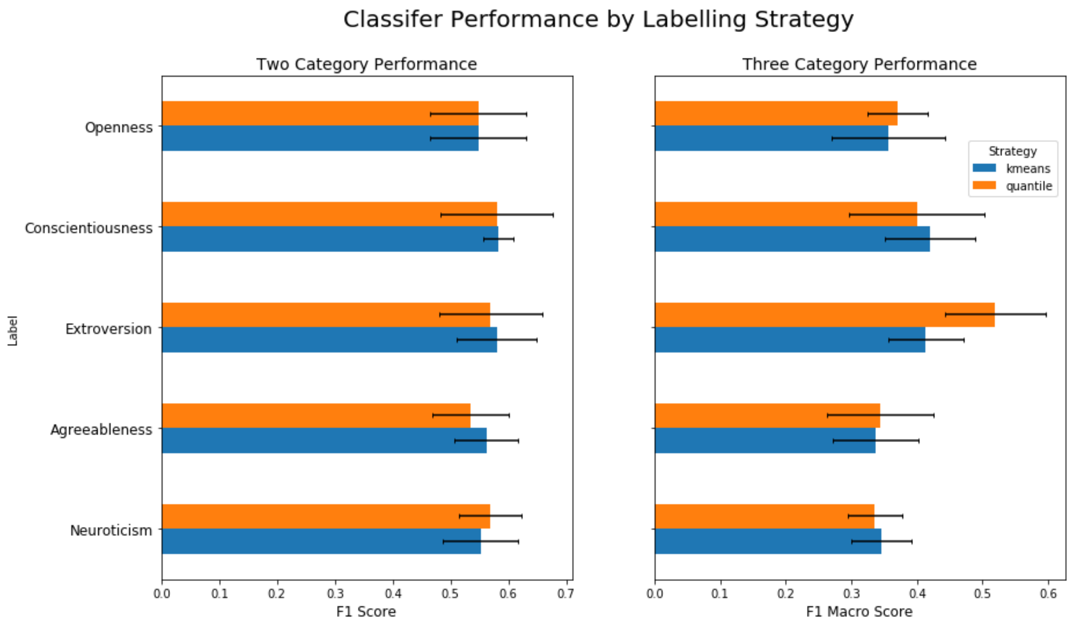


FIGURE 6.5: Classifier Performance by Labelling Strategy.
Error bars: Standard Deviation

(Ojala & Garriga, 2010). We correct all p-values for multiple comparisons using the Benjamini-Hochburg procedure (Benjamini & Hochberg, 1995). To allow comparison to previous studies, we still report the accuracy of classifiers.

### 6.5.2 Results

Across the feature groups, the traits of Conscientiousness and Extroversion were predicted significantly better than chance when using the k-means based categories. In comparison, only Extroversion was predicted significantly better than chance using the quantile-based categories. We find that, despite the different labelling strategies, the same algorithm and feature groups lead to significantly better performance than chance for Extroversion (Feature Group: Eye Movement Statistics, Algorithm: Ridge). No feature group or labelling strategy gave significant insight into the personality traits of Openness, Agreeableness or Neuroticism. We provide a comparison of the $F1_{macro}$ scores across all five traits in Figure 6.5, which shows the best $F1_{macro}$ score achieved for each trait in the three-category instances, and the best $F1$ score achieved for each trait in the two-category instances.

To evaluate whether there was a significant effect of strategy, we performed a

TABLE 6.3: Three Category Classification Results for Conscientiousness and Extroversion by Labelling Strategy

| Trait | Feature Group (Algorithm) | Accuracy Mean (SD) | $F1_{Macro}$ Mean (SD) | Baseline Accuracy |
|---|---|---|---|---|
| **K-Means Labelling Strategy** | | | | |
| Conscientiousness* | AOI Prop+ (Ridge) | 50.6% (8.7) | .419 (.069) | 42.0% |
| Extroversion* | EMS (Ridge) | 46.1% (6.7) | .413 (.058) | 43.4% |
| **Quantile Labelling Strategy** | | | | |
| Conscientiousness† | AOI Prop (Ridge) | 42.8% (8.5) | .398 (.079) | 36.1% |
| Extroversion*** | EMS (Ridge) | 53.3% (7.3) | .519 (.077) | 40.6% |

* p<.05. Ridge: One-vs-Rest Ridge Classification. k-NN: K-nearest neighbors.
+ indicates with frequency metrics. † uncorrected p-value= .036.

repeated measures t-test upon the outer k-fold $F1_{macro}$ scores across all five personality traits (five traits, five folds = 25 scores per strategy) for both the two and three-category conditions. As scores for the k-means strategy in the two-category condition violated the assumption of normality (W(24)= 0.917, p=.043), we report the Wilcoxon signed rank test results. We found no evidence of a significant effect of the labelling strategy across the five traits for either the two (W(49) = 104, p=.970) or three (t(49) = -0.92, p=.365) category conditions.

When considering just trait Conscientiousness and Extroversion, visual inspection of Figure 6.5 indicates that, while there is no clear difference in classifier performance in the two-category instances, in the three-category instances there does appear to be an influence of the labelling strategy. We highlight this in Table 6.3. This illustrates that, in the three-category classification situation, the *k*-means based labels lead to slightly better $F1_{macro}$ scores for Conscientiousness, and to considerably worse $F1_{macro}$ scores for Extroversion.

### 6.5.3 Interpreting Classifier Results

Our findings do not support our hypothesis and suggest that increasing the silhouette score of the personality trait categories does not make a significant impact upon the final performance of the classifiers. While overall our findings suggest that neither strategy has a significant advantage over the other, we note that performance varies for certain traits such as Extroversion and Conscientiousness in the three-category condition, where it appears that the quantile strategy leads to better performance for Extroversion, and worse performance for Conscientiousness. As such, we recommend carefully exploring the effect of the clustering strategy upon silhouette scores (and closely evaluating the appropriateness of the approach via visualising the dendrogram) before applying the k-means approach as it may lead to reduced classifier performance.

## 6.6 Study Three

In a recent review of why people use Facebook, it was identified that two main motivating factors are a need to belong and a need for self-presentation (Nadkarni & Hofmann, 2012). This forms a two-factor model of Facebook use whereby demographic and psychological factors modulate how and why users engage with Facebook. As visual behaviour is driven in part by what we find rewarding in the scene (Hayhoe & Ballard, 2005), we propose that the same demographic and psychological factors that modulate Facebook behaviour may also be reflected in the individual's visual behaviour. As we have already established that personality traits such as Extroversion and Conscientiousness can be predicted from visual behaviour, we investigate novel attributes that are linked to Facebook use to understand if they may be predictable from visual behaviour upon Facebook.

Previous literature has suggested that an individual's need for self-presentation (e.g., wanting to be perceived positively by others) is related to both their trait narcissism score and self-esteem (Andreassen, Pallesen, & Griffiths, 2017). Further to this, both narcissism and self-esteem have been found to correlate with how and why the individual engages with Facebook (Andreassen et al., 2017; Brailovskaia & Margraf, 2018; Nadkarni & Hofmann, 2012; Ryan & Xenos, 2011) and thus may be attributes that can be predicted from Facebook use. In particular, narcissism is characterised by a greater interest in the symbolic than utilitarian value of products, leading to the hypothesis that the more narcissistic someone is, the more they will spend more time fixating pictorial rather than linguistic information (Cisek et al., 2014). Furthermore, there is evidence to suggest that narcissistic individual's have specific patterns of neurological connectivity within reward evaluation circuits, which may lead to this behaviour (Chester, Lynam, Powell, & DeWall, 2016). As visual behaviour is guided in part by what the individual finds rewarding in the visual scene, this provides a functional account of how narcissism (seeking symbolic value and pictorial stimulus) and self-esteem (seeking positive evaluation) may lead to distinct differences in the attribution of visual attention while viewing Facebook content. This leads us to the hypothesis that the traits of narcissism and self-esteem may be reflected in, and thus be predictable from, eye movements upon Facebook content.

Narcissism and self-esteem are not the only personal attributes that may influence social media use; previous literature has been able to predict an individual's sex significantly better than chance from records of online SNS use (Kosinski et al., 2013), and from eye movements in visual search tasks (Andersen, Dahmani, Konishi, & Bohbot, 2012; Bargary et al., 2017). Therefore, sex may also be reflected in eye movements upon social media site content. Furthermore, the personal attributes of political inclination and whether an individual voted have previously been predicted from records of online SNS use (Kosinski et al., 2013), but not from their visual behaviour.

In previous studies, we have shown that aspects of an individual's personality can be predicted from their eye movements whilst browsing Facebook content. Now, in this study, we wish to identify if attributes outside of personality can also be predicted from this visual behaviour. Therefore, in a broad exploration of personal attributes outside of personality, we evaluate whether an individual's narcissism category, self-esteem, political inclination (including whether they voted), and sex can be predicted from twenty seconds of visual behaviour while browsing their own Facebook NewsFeed. This expands the literature by identifying whether attributes linked to social media use are also reflected in eye movements during social media use, which provides a broader picture of the potential privacy risks associated with sharing eye movement data with SNS companies.

### 6.6.1 Methods

Using the feature groups outlined in Section 6.1, and the machine learning methodology outlined in Section 6.5.1, we report all models that perform significantly better than chance ($F1$ based p-values calculated via permutation tests). As part of our previous study, participants responded to a series of demographic questions and questionnaires which were not part of the main hypothesis. We outline the additional materials in the following sections. For the outcomes of Sex and Vote, we directly predict the existing categorical structure (e.g., ('Male'/'Female', 'No'/'Yes'). For outcomes that do not have a clear category structure (e.g., questionnaire scores) we carefully choose our labelling strategy by following the outlined procedure in study one, and select the k-means clustering approach only when strongly indicated by the data (e.g., there is a clear separation with non-overlapping confidence intervals between the techniques when evaluating silhouette scores). The visual behaviour collected is described in section 6.3.2.

**Narcissism Questionnaire**

Participants completed the NPI-16 short measure of narcissism (Cronbach's alpha .72; Ames, Rose, & Anderson, 2006), consisting of sixteen two-alternative forced choice questions where the participant must select the answer they agree with most (e.g., 'I prefer to blend in with the crowd' or 'I like to be the center of attention'). One answer for each question is narcissism-congruent, and the participant receives one point per narcissism-congruent answer. The trait narcissism score is calculated as the mean across the sixteen items, with a minimum score of zero (no narcissism-congruent answers selected) and a maximum of one (all narcissism-congruent answers selected).

**Self-Esteem Questionnaire**

Participants completed the ten-item Rosenberg self-esteem scale (Cronbach's alpha 0.89; Robins, Hendin, & Trzesniewski, 2001). Participants respond to each item upon a four-point Likert scale (1-4) ranging from 'Strongly Disagree' to 'Strongly Agree'. Five of the ten items are reverse scored (e.g., a reverse scored item is 'At times I think I am no good at all'). The participant's self-esteem is calculated as the sum of the questionnaire items (after correcting for reverse scoring), with a maximum of 40 and a minimum of 10.

#### Voting Behaviour and Preferences

Participants were asked if they voted in the last election ('Yes', 'No') and reported their political inclination upon a hundred point visual analogue slider, centered at zero and spanning the range -1 (Extremely Liberal) to +1 (Extremely Conservative).

### 6.6.2 Results

#### Descriptive Statistics: Binary Outcomes

We have two binary outcomes, sex ('Male / Female') and vote ('Yes' / 'No'). Vote refers to whether the participant cast a vote within the 'last election'. While this can generally be interpreted as the 2019 United Kingdom (UK) General Election, we note that the young international cohort leads to ambiguity in a negative response as they may not have been able to vote. Seven individual's did not respond to the 'Vote' question (N=173). There are 107 who did not vote and 66 who did. For the sex outcome, we note a large class imbalance due to our mostly female cohort (144 female, 36 male, no missing responses). To address this, we randomly sampled 36 females from the cohort to match the 36 males.

#### Descriptive Statistics: Questionnaire Based Outcomes

For the questionnaire based outcomes there were no missing values. We employ the same evaluation procedure for deciding the number of categories as in section 6.4. Interpreting the dendrogram for political inclination suggests that two categories best reflected the underlying data, and we employ a quantile-split approach as there was no clear advantage of employing the *k*-means strategy upon silhouette scores for the two-category case. For self-esteem, the dendrogram structure suggested that the three categories may best reflect the underlying data, and we employ a *k*-means strategy as there was a clear advantage in the silhouette score for the three-class instance. For narcissism, the dendrogram indicated that either three classes would represent the underlying data. There was a non-overlapping advantage for using the *k*-means labels, but we note that the solution resulted in a very strong class imbalance ([Class One: 121, Class Two: 46, Class Three: 13]). Machine learning techniques learn from experience, and only thirteen observations being in the third category

TABLE 6.4: Descriptive Statistics for Continuous Personal Attributes Assigned to Categories

| Attribute | Strategy | Mean (SD) | | | Support |
|---|---|---|---|---|---|
| | | Low | Medium | High | |
| Self Esteem | $k$-means | 10.64 (2.18) | 16.57 (1.52) | 22.43 (2.59) | [42, 108, 30] |
| Narcissism | Quantile | .027 (.031) | .157 (.032) | .399 (.156) | [46, 53, 81] |
| Political Inclination | Quantile | -.630 (.188) | - | -.034 (.199) | [89, 91] |

TABLE 6.5: Significant Results for Personal Attributes.

| Label | Feature Group | Algorithm | Accuracy Mean (SD) | F1 / F1 Macro Mean (SD) | Categories |
|---|---|---|---|---|---|
| Narcissism* | EMS | Ridge | 47.8% (5.7) | 0.4254 (.052) | 3 |
| Narcissism* | Page Content | k-NN | 46.7% (5.9) | 0.4156 (.038) | 3 |
| Self Esteem* | EMS | Naive Bayes | 57.2% (11.3) | 0.4357 (.098) | 3 |
| Sex* | EMS | SVM | 63.0% (16.0) | 0.6433 (.184) | 2 |
| Sex* | EMS | Ridge | 62.9% (11.6) | 0.6286 (.132) | 2 |
| Sex* | AOI+ | Ridge | 62.7% (10.6) | 0.6385 (.105) | 2 |

* p<.05. + indicates with frequency metrics.
Ridge: One-vs-Rest Ridge Classification. k-NN: K-nearest neighbors. SVM: Support vector machine.

are highly likely to be too few to learn from. Furthermore, as outlined, our scoring scheme will heavily penalise a classifier if it struggles with any one category. As such, we make the decision to employ the quantile based technique. We provide an outline of the number of categories and the splitting strategies chosen for each personal attribute (including sex and vote) in Table 6.4.

**Classifier Results**

Out of our five personal attributes explored, we find that an individual's narcissism and self-esteem category ('low', 'medium', 'high') along with their sex ('female', 'male') can be predicted significantly better than expected from chance (Table 6.5) based upon $F1$ score. For narcissism, both descriptions of visual behaviour and the page content alone were significantly informative. For self-esteem, only descriptions of visual behaviour were informative. For sex, only descriptions of visual behaviour were informative. The best performing classifiers for both political inclination ($F1 = .583$, $Accuracy = 52.5\%$) and Vote ($F1 = .523$, $Accuracy = 45.7\%$) were not significantly better than chance.

### 6.6.3 Interpreting Classifier Results for Novel Attributes

Further from the already identified Extroversion and Conscientiousness, we proposed that the same personal attributes that influence how an individual interacts with Facebook may also be reflected in their eye movements upon the social media platform. We find partial support for this suggestion, with the attributes of sex, narcissism and self-esteem being predicted significantly better than chance from twenty seconds of viewing behaviour upon their own Facebook NewsFeed.

#### Narcissism

When predicting narcissism from visual behaviour, statistical descriptions (i.e., of fixation and saccadic properties) were more informative than the AOI based metrics (e.g., describing visual behaviour in response to text or image based content). We also found that the type of content present within an individual's Facebook page is almost equally as informative of their narcissism category as statistical descriptions of eye movements are.

Given that a *k*-neighbour algorithm produces the best performance when predicting narcissism from page content information, we can deduce that individuals with similar narcissism scores have similar distributions of content types (e.g., images, text or video) upon their Facebook NewsFeed. Therefore it may be that narcissism, due to its link to reward systems (Chester et al., 2016), influences the user's past behaviour and prompts the Facebook recommender system to present a distinctive distribution of content types, which leads to page content information being informative of trait Narcissism. On average, participants viewed roughly two to four posts during their twenty seconds of recorded viewing duration, which suggests that this distribution emerges within the first four or so pieces (Facebook 'Posts') of content. It is also possible that having similar content leads to a similar distribution of eye movements, which our algorithms may have picked up upon when being trained upon statistical eye metrics.

While our results hint that narcissism may be related to a distinct pattern of online behaviour (that leads to similar content being shown in the NewsFeed), it does not appear that this manifests in how the individual visually explores different content types (as hypothesised by Cisek et al., 2014). This is because a distinct pattern of spending more time fixating pictorial rather than linguistic information would lead to the area of interest (rather than content-agnostic eye movement statistics) metrics being the most informative. Due to the compromise between predictive and exploratory power employed within this study further work is required to elucidate the relationship between narcissism and eye movements upon social network based stimuli, with potential questions including whether the association between narcissism and visual behaviour is highly context-dependent (e.g., is closely linked to social media use) or generalises across different domains.

#### Self Esteem

We identify that self esteem can be predicted significantly better than chance from statistical descriptions of eye movements. This supports previous research that found a link between self esteem and Facebook use (Andreassen et al., 2017), and suggests that, in a similar way to narcissism, self esteem may lead to distinct patterns of oculomotor events (fixations and saccades). Interestingly, in contrast to narcissism, it appears that self-esteem is more related to how the participant explores the page rather than the types of content upon the page (as page content information was not informative). It is possible that this association is driven by the participant's seeking of self-validating material in the visual environment, and future research may wish to investigate this further within experimental settings.

#### Sex

We find that an individual's sex can be predicted significantly better than chance from twenty seconds of visual behaviour upon their own Facebook NewsFeed. This finding is consistent with previous research that has illustrated that there are sex differences in visual behaviour across a range of visual tasks (Andersen et al., 2012; Bargary et al., 2017), and that sex is predictable from records of online behaviour upon Facebook (Kosinski et al., 2013). Our finding is novel in two aspects, the first being the short time period, and the second being the ecologically valid domain. Previous literature predicting sex from visual behaviour has focused upon controlled experimental conditions (Bargary et al., 2017) which do not always generalise to real-world conditions (Tatler, Hayhoe, Land, & Ballard, 2011). As such, our finding illustrates that sex differences in the distribution of visual attention also emerge within the ecologically valid setting of the user's own Facebook NewsFeed, and can be used to distinguish males from females significantly better than expected by chance. Similarly, previous literature predicting sex from online behaviour does so over records spanning a much greater period of time (Settanni, Azucar, & Marengo, 2018), and we illustrate here that twenty seconds is enough to achieve above-chance performance. Regarding which descriptions of visual behaviour give insight into an individual's sex, it appears that both content-based and statistical (content-agnostic) descriptions of visual behaviour are informative. This suggests that there is a relationship between sex and the statistical properties of the eye movements (e.g., Fixations and saccades, as found in: Bargary et al., 2017), and that there is a distinct pattern of attentional allocation to different types of content within the visual scene.

**Other Attributes**

Previous research has found that an individual's political inclination and whether they voted in the last election can be predicted from the distribution of their 'likes' upon Facebook (Kosinski et al., 2013). In our paradigm, we do not find these attributes to be predictable from visual behaviour upon Facebook. While it is possible that this information is simply not present within the data that was collected, we highlight three possible limitations to this interpretation. Firstly, there are key differences in the political party system between our UK cohort and the American bipartisan system that may hinder direct comparison, and future research may wish to replicate this study within an American cohort. Secondly, we did not encode the content upon a fine-grained scale (e.g., we did not label whether the content was political propaganda or not). It is possible that describing visual behaviour in response to categories of content that reflect political information (e.g., pro or anti-brexit) may reflect aspects of the individual's political inclination. This is outside the scope of the current study, but provides an avenue for future research. Finally, our cohort is heavily left-leaning, and in the average case we may be trying to separate strongly liberal versus centrist individuals based upon their eye movements rather than liberal versus conservative (see mean values for each political category in Table 6.4). As such, future research may wish to address the above concerns within a balanced cohort, where both liberal and conservative viewpoints are equally represented.

## 6.7 Overall Discussion

The application of machine learning principles to psychological questions is becoming widespread and provides several benefits over traditional approaches within the field. Namely, the focus of the method upon accurate prediction, at the cost of interpretability, allows the researcher to understand the model's ability to generalise to previously unseen observations (Yarkoni & Westfall, 2017). Often, to apply machine learning algorithms that are incompatible with the regression problem, continuous outcomes such as an individual's trait score are segmented into categories (e.g., low, medium and high as in: Berkovsky et al., 2019; Hoppe et al., 2018). In this paper, we explored how unsupervised learning techniques may help researchers form more appropriate categories than those formed from the currently employed median split, or other quantile-based binning strategies. Using the silhouette score, we evaluated the quality of the categories produced by each strategy. The silhouette score reflects both how similar items within each category are, and how distinct items across categories were. We found that using a *k*-means clustering algorithm, with the number of clusters guided via visual inspection of the dendrogram from agglomerative clustering, as expected led to categories that equaled or exceeded the cluster quality of the quantile based strategy. The main benefit of our suggested approach is that it makes fewer assumptions about the underlying data, is able to provide insight into the underlying structure of the data, and can inform researchers upon likely candidates for the number of categories to form. Especially important is that the methodology informs the researcher on the number of categories to form for each trait independently, which is an advancement upon arbitrarily splitting all traits into two or three categories as has been conducted in previous literature. However, when testing whether this splitting strategy leads to improved performance when predicting personality traits from visual behaviour in study two, in contrast to our hypothesis, we found no significant effect of the splitting strategy upon the best performing classifier's $F1_{macro}$ score across the five personality traits. This suggests that increasing the silhouette score of the personality trait categories does not always make a significant impact upon the final performance of the classifiers. The splitting strategy employed did, however, influence the classifier performance for trait Extroversion in the three-class instance (see Table 6.3), demonstrating that the labelling strategy employed can lead to different classification results. As such, we propose that the best practice is to apply our two step approach only when it demonstrates a substantial advantage in cluster cohesion over the quantile-split technique, and consider category formation upon a case-by-case basis as illustrated in study three. Importantly, within the analysis of study two, we have restrained ourselves to forming categories upon the calculated trait scores as defined within (McCrae & Costa, 2004), as opposed to the trait-congruent questionnaire items. This gives a single score for each participant and constrains the clustering problem to finding the range each category should span upon the single trait dimension. This is directly equivalent to the problem solved by quantile based strategies within previous research (Berkovsky et al., 2019; Hoppe et al., 2018). The advantage of this approach is that categories are readily interpreted as being upon the continuum of low to high (e.g., each category directly maps on to the overall trait concept). However, in contrast to the quantile split technique, the two-step clustering approach presented in this paper generalises to an arbitrary number of dimensions. As such, future research may wish to investigate employing our technique upon the questionnaire item level responses. Each personality trait is composed of twelve questions, which are further subdivided into components of the trait. For example, Conscientiousness is composed of Competence, Order, Dutifulness, Achievement Striving, Self-Discipline and Deliberation (Moutafi, Furnham, & Crump, 2006). The advantage of clustering upon the questionnaire responses is that it allows the technique to pick up on systematic patterns of responses within each item, or component, that might otherwise be lost during the aggregation of the trait score. The disadvantage is that this may result in clusters that have no easily interpreted theme and may be less theoretically constrained than the approach taken within this paper. A direction for future research would be to explore this theme, to determine whether the solutions formed from such an approach lead to reasonable categories, and whether these categories improve classifier performance. However, as the silhouette score only indicates the goodness of fit, rather than the theoretical validity of the categorisation strategy, without an objective function by which to assess theoretical validity, it would be difficult to evaluate such clustering solutions. A possible approach would be to cross-validate the theoretical validity of the formed categories with external measures that compliment the self-reported personality questionnaire data, such as gaining observational measures of participant behaviour (Oh, Wang, & Mount, 2011).

Finally, in our third study we utilised the framework developed in experiments one and two to predict personal attributes from visual behaviour that were not part of the main study in Chapter 5. We investigated the attributes of narcissism, self-esteem, political inclination,voting, and sex. In a novel contribution to the literature, we identified that an individual's sex, self-esteem, and narcissism categories can be predicted significantly better than chance from twenty seconds of visual behaviour upon their own Facebook NewsFeed. This illustrates that a wide range of attributes, such as an individual's personality characteristics (conscientiousness and extroversion), attributes (self-esteem and narcissism), and demographic factors (sex) are reflected within oculomotor behaviour upon social media type content.

### 6.7.1 Generalising beyond current results: limitations and considerations.

As with all studies, it is important to understand to which population our results can be generalised to. We note that our findings relate to a young, mainly female cohort. It is possible that our young cohort (Mean age: 20.5) may restrict the generalisability of our findings, with younger adults being associated with more frequent Facebook use with less of a focus upon political or current affairs topics (Holt, Shehata, Strömbäck, & Ljungberg, 2013). However, an investigation into the role of narcissism in the type of gratification sought upon Facebook found no generational differences (Leung, 2013), which suggests that the generalisability of our results relating to narcissism and self-esteem may not be impacted by the mean age of our cohort.

Perhaps more important is the gender imbalance in our cohort. In the general UK population, the ratio of males to females is roughly equal (Swan, 2012), and we note that only 20% of our cohort were male. As shown in study three, sex is reflected in visual behaviour, and the underlying distribution of males to females in our study is imbalanced. This may provide a limitation when generalising our findings outside of female-biased samples, as previous research highlights gender differences in the motivations for why individuals use Facebook (Andreassen et al., 2017). As such, future research may wish to ensure that they sample from a more balanced cohort. We note that this does not influence the interpretation of the results related to predicting sex from visual behaviour in study three, as we randomly resampled to gain an equal number of females and males (see Table 6.4). However, compared to the previous evaluations, this represents a relatively small sample (72 instances versus 180) and may also wish to be replicated in a larger cohort.

### 6.7.2 Conclusion

In conclusion, we have proposed and evaluated an alternative to quantile binning when splitting continuous trait personality scores into discrete categories. The advantage of the new approach is that fewer assumptions are made about the underlying data, and the technique provides an insight towards selecting an appropriate number of categories. We wish to highlight that, in data driven designs, the optimal number of clusters is driven by two key factors. The first is a matter of practicality regarding the number of participants available, as there must be a suitable amount of observations in each cluster for the algorithm to learn from. As such, a small sample size tends to preclude the selection of a large number of clusters. The second factor is the natural structure inherent within the data, which is visualised in the agglomerative clustering plots. A key contribution of the proposed methodology is to allow the researcher to make informed decisions about the trade-off between these two factors by selecting at which level of agglomeration (i.e., where on the y-axis of the dendrogram) provides the most appropriate categorical split for their cohort. These considerations also impact the theoretical contributions of this paper, as our selection of the number of clusters in this study is influenced both by our relatively small sample size (under 200 participants) and the inherent structure found within our sample. Because of this, it is important to note that different studies with different sample sizes (or cohort characteristics) may find alternative solutions are optimal for their given application. For example, a study with tens of thousands of participants may well wish to categorise their participants into a larger number of clusters to achieve a higher level granularity; the main theoretical constraint is that the clusters are sufficiently distinct as to be meaningful, which can only be interpreted in light of the chosen application.

In this study, we demonstrated that our proposed technique leads to categories that exhibit the properties of being more cohesive and better distinguished than the commonly employed quantile split technique. A key limitation of our suggested approach is that it does not reliably lead to significant differences in classifier performance across the five personality traits investigated, although we observe that it does lead to varying classifier performance within specific personality traits. As such, overall we propose that the best practice is to evaluate the evidence provided by our two-step clustering procedure and select the k-means clustering approach only when strongly indicated by the data (e.g., there is a clear separation with non-overlapping confidence intervals between the *k*-means and quantile split techniques when evaluating silhouette scores).

In study three, we employed a practical test of this methodology by investigating if novel personal attributes can be predicted from visual behaviour upon Facebook. In a novel contribution to the literature, we find that sex, self-esteem and narcissism can be predicted from twenty seconds of visual behaviour upon the participants own Facebook NewsFeed. This analysis highlights that a wide range of personal attributes can be predicted from visual behaviour, and perhaps most importantly, supports the proposition that any psychological factor linked to how or why the participant engages with social media content may be a promising candidate to predict from visual behaviour. We suggest that future research may wish to replicate our findings within a gender-balanced cohort, and explore how personal attributes such as narcissism influence the search pattern of individual's across various domain types (e.g., social media content, classical visual search tasks) to better characterise this association. Finally, we note that the application of advanced pattern recognition techniques to visual behaviour provides a new frontier in human-computer interaction and promote the awareness that with these new advances come new security and privacy concerns.

## References


Alpaydin, E. (2014). *Introduction to machine learning*. MIT press.

Ames, D. R., Rose, P., & Anderson, C. P. (2006). The NPI-16 as a short measure of narcissism. *J. Res. Pers. 40*(4), 440–450. doi:10.1016/j.jrp.2005.03.002

Andersen, N. E., Dahmani, L., Konishi, K., & Bohbot, V. D. (2012). Eye tracking, strategies, and sex differences in virtual navigation. *Neurobiology of Learning and Memory*, *97*(1), 81–89. doi:10.1016/j.nlm.2011.09.007

Andreassen, C. S., Pallesen, S., & Griffiths, M. D. (2017). The relationship between addictive use of social media, narcissism, and self-esteem: Findings from a large national survey. *Addictive Behaviors*, *64*, 287–293. doi:10.1016/j.addbeh.2016.03.006

Arnoux, P.-H., Xu, A., Boyette, N., Mahmud, J., Akkiraju, R., & Sinha, V. (2017). 25 Tweets to Know You: A New Model to Predict Personality with Social Media. In *AAAI Conference on Web and Social Media* (Vol. 11, pp. 472–476). Montreal, Canada: AAAI Press.

Baranes, A., Oudeyer, P. Y., & Gottlieb, J. (2015). Eye movements reveal epistemic curiosity in human observers. *Vision Res. 117*, 81–90. doi:10.1016/j.visres.2015.10.009

Bargary, G., Bosten, J. M., Goodbourn, P. T., Lawrance-Owen, A. J., Hogg, R. E., & Mollon, J. (2017). Individual differences in human eye movements: An oculomotor signature? *Vision Res. 141*, 157–169. doi:10.1016/j.visres.2017.03.001

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, *57*(1), 289–300.

Berkovsky, S., Taib, R., Koprinska, I., Wang, E., Zeng, Y., Li, J., & Kleitman, S. (2019). Detecting Personality Traits Using Eye-Tracking Data. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (pp. 1–12). CHI '19. doi:10.1145/3290605.3300451

Brailovskaia, J., & Margraf, J. (2018). What does media use reveal about personality and mental health? An exploratory investigation among German students. *PLoS One*, *13*(1). doi:10.1371/journal.pone.0191810

Chester, D. S., Lynam, D. R., Powell, D. K., & DeWall, N. (2016). Narcissism is associated with weakened frontostriatal connectivity: A DTI study. *Soc. Cogn. Affect. Neurosci. 11*(7), 1036–1040. doi:10.1093/scan/nsv069

Cisek, S. Z., Sedikides, C., Hart, C. M., Godwin, H. J., Benson, V., & Liversedge, S. P. (2014). Narcissism and consumer behaviour: A review and preliminary findings. *Front. Psychol. 5*, 232. doi:10.3389/fpsyg.2014.00232

Eftekhar, A., Fullwood, C., & Morris, N. (2014). Capturing personality from Facebook photos and photo-related activities: How much exposure do you need? *Comput. Human Behav. 37*, 162–170. doi:10.1016/j.chb.2014.04.048

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*. ROC Analysis in Pattern Recognition, *27*(8), 861–874. doi:10.1016/j.patrec.2005.10.010

Ferwerda, B., Schedl, M., & Tkalcic, M. (2015). Predicting Personality Traits with Instagram Pictures. In *Proc. 3rd Work. Emot. Personal. Pers. Syst. 2015 - Emp. '15* (pp. 7–10). doi:10.1145/2809643.2809644

Hayhoe, M., & Ballard, D. (2005). Eye movements in natural behavior. *Trends Cogn. Sci. 9*(4), 188–194. doi:10.1016/j.tics.2005.02.009

Holt, K., Shehata, A., Strömbäck, J., & Ljungberg, E. (2013). Age and the effects of news media attention and social media use on political interest and participation: Do social media function as leveller? *European Journal of Communication*, *28*(1), 19–34. doi:10.1177/0267323112465369

Hoppe, S., Loetscher, T., Morey, S. A., & Bulling, A. (2018). Eye Movements During Everyday Behavior Predict Personality Traits. *Front. Hum. Neurosci. 12*, 105. doi:10.3389/FNHUM.2018.00105

Kosinski, M., Stillwell, D., & Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. *Proc. Natl. Acad. Sci. 110*(15), 5802–5805. doi:10.1073/pnas.1218772110

Leung, L. (2013). Generational differences in content generation in social media: The roles of the gratifications sought and of narcissism. *Computers in Human Behavior*, *29*(3), 997–1006. doi:10.1016/j.chb.2012.12.028

McCrae, R., & Costa, P. (2004). A contemplated revision of the NEO Five-Factor Inventory. *Pers. Individ. Dif. 36*(3), 587–596. doi:10.1016/S0191-8869(03)00118-1

Mirkin, B. (2011). Choosing the number of clusters. *WIREs Data Mining and Knowledge Discovery*, *1*(3), 252–260. doi:10.1002/widm.15

Moutafi, J., Furnham, A., & Crump, J. (2006). What facets of openness and conscientiousness predict fluid intelligence score? *Learning and Individual Differences*, *16*(1), 31–42. doi:10.1016/j.lindif.2005.06.003

Nadkarni, A., & Hofmann, S. G. (2012). Why do people use Facebook? *Pers. Individ. Dif. 52*(3), 243–249. doi:10.1016/J.PAID.2011.11.007

Nummenmaa, L., Hyönä, J., & Calvo, M. G. (2006). Eye movement assessment of selective attentional capture by emotional pictures. *Emotion*, *6*(2), 257–268. doi:10.1037/1528-3542.6.2.257

Oh, I.-S., Wang, G., & Mount, M. K. (2011). Validity of observer ratings of the five-factor model of personality traits: A meta-analysis. *Journal of Applied Psychology*, *96*(4), 762–773. doi:10.1037/a0021832

Ojala, M., & Garriga, G. C. (2010). Permutation Tests for Studying Classifier Performance. *J. Mach. Learn. Res. 11*(Jun), 1833–1863.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res. 12*(Oct), 2825–2830.

Rauthmann, J. F., Seubert, C. T., Sachse, P., & Furtner, M. R. (2012). Eyes as windows to the soul: Gazing behavior is related to personality. *J. Res. Pers. 46*(2), 147–156. doi:10.1016/j.jrp.2011.12.010

Robins, R. W., Hendin, H. M., & Trzesniewski, K. H. (2001). Measuring global self-esteem: Construct validation of a single-item measure and the Rosenberg Self-Esteem Scale. *Personal. Soc. Psychol. Bull. 27*(2), 151–161. doi:10.1177/0146167201272002

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, *20*, 53–65. doi:10.1016/0377-0427(87)90125-7

Ryan, T., & Xenos, S. (2011). Who uses Facebook? An investigation into the relationship between the Big Five, shyness, narcissism, loneliness, and Facebook usage. In *Comput. Human Behav.* (Vol. 27, pp. 1658–1664). doi:10.1016/j.chb.2011.02.004

Segalin, C., Celli, F., Polonio, L., Kosinski, M., Stillwell, D., Sebe, N., … Lepri, B. (2017). What your Facebook Profile Picture Reveals about your Personality. doi:10.1145/3123266.3123331

Settanni, M., Azucar, D., & Marengo, D. (2018). Predicting Individual Characteristics from Digital Traces on Social Media: A Meta-Analysis. *Cyberpsychology, Behav. Soc. Netw. 21*(4), 217–228. doi:10.1089/cyber.2017.0384

Shimojo, S., Simion, C., Shimojo, E., & Scheier, C. (2003). Gaze bias both reflects and influences preference. *Nat. Neurosci. 6*(12), 1317–1322. doi:10.1038/nn1150

Steinley, D., & Brusco, M. J. (2011). Choosing the number of clusters in *K*-means clustering. *Psychological Methods*, *16*(3), 285–297. doi:10.1037/a0023346

Swan, J. (2012). *2011 Census: Population Estimates for the United Kingdom*. Office for National Statistics.

Tatler, B., Hayhoe, M., Land, M., & Ballard, D. (2011). Eye guidance in natural vision: Reinterpreting salience. *J. Vis. 11*(5), 5–5. doi:10.1167/11.5.5

Tatler, B. W., Wade, N. J., Kwan, H., Findlay, J. M., & Velichkovsky, B. M. (2010). Yarbus, eye movements, and vision. *Iperception. 1*(1), 7–27. doi:10.1068/i0382

Ward, J. H. (1963). Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association*, *58*(301), 236–244. doi:10.1080/01621459.1963.10500845

Yarkoni, T., & Westfall, J. (2017). Choosing Prediction Over Explanation in Psychology: Lessons From Machine Learning. *Perspect. Psychol. Sci. 12*(6), 1100–1122. doi:10.1177/1745691617693393

# Chapter 7

# Discussion

In the introduction, I discussed how eye movements are influenced by much more than just the physical properties of the scene, and how they may reflect aspects of the individual's personal attributes. This provides the rationale behind the main goal of this thesis, which sets out to establish whether personal attributes may be predicted from the eye movements of users while they browse online social networking site (SNS) content. In the following sections, I outline how the research included within this thesis has addressed the main research objectives, reflect on what gaps remain in the theory put forward, and provide suggestions for future research.

## 7.1 Contributions to understanding the risk of unknowingly disclosing personal information via eye tracking technology

The first empirical study investigated whether the same associations found between visual behaviour and preference decisions within static image paradigms (e.g., as found by: Holmes & Zanker, 2012; Nummenmaa, Hyönä, & Calvo, 2006; Shimojo, Simion, Shimojo, & Scheier, 2003) replicate when participants browse within a novel social media style stimulus (Chapter 3). The results from this study illustrate that a significant amount of variance in our participants affective evaluations of the stimuli (Facebook posts) was explained by a linear combination of three oculomotor metrics (time to first fixation, total fixation duration and number of fixations). This provides the first piece of evidence from this thesis supporting the claim that eye movements within naturalistic stimuli are significantly associated with personal information, in this case a subjective evaluation. A surprising pattern of associations was found between our participants viewing behaviour and their affective evaluation of the stimuli, compared to that found within previous literature. Namely, our study found that the number of fixations was positively correlated with the participant's rating, yet the total fixation duration was negatively associated with the participant's rating; in previous studies, both fixation duration and number have been found to be positively associated, as one might expect (e.g., Holmes & Zanker, 2012; Nummenmaa et al., 2006). This was thought to occur primarily due to the different visual exploration paradigm offered by social media style web sites, which, in contrast to previous literature, contained text and allowed the participant to scroll to a new stimulus at will.

While statistically significant, the resultant model based upon multiple linear regression explained very little variance (5%) in the outcome variable of emotional rating (from -1 'negative' to +1 'positive') in this study. In an exploratory analysis, this was increased to 10% by seeking to predict the absolute (i.e., ranging from 0 'low' to 1 'high') emotional rating, yet this is still a very weak association. As such, while having important implications for those conducting market research via eye tracking techniques (covered in detail within the paper), in regard to the main topic of the thesis, this result indicates that the association between an individual's preferences and eye movement behaviour manifests as only a very weak linear correlation. As such, my preliminary finding is that eye movements upon SNS content reflects cognitive appraisal. However, the multiple linear regression models formed in this study pose little practical risk to users privacy; a user's subjective preferences are unlikely to be accurately predicted from their visual behaviour upon social media style content using this technique.

To ensure that the results were directly comparable to previous literature (e.g., Holmes & Zanker, 2012; Maughan, Gutnikov, & Stevens, 2007; Nummenmaa et al., 2006), this study explored the linear associations between a limited set of variables (i.e., three oculomotor metrics) and the outcome variable. Of note is that the outcome variable itself demonstrated ceiling effects, with many participants giving the maximum rating of positive one (see figure: 3.4). This may have led to our results underestimating the association between our three metrics and emotional valence, as it violates the assumption of linearity. Namely, this ceiling effect means that the interaction between the dependent variables (oculomotor metrics) and the outcome variable becomes nonlinear after the maximum score as it can no longer increase, thus, it is not able to be captured by linear methods. In the context of this thesis, a lesson from this study is that it is important to ensure that the personal attribute being measured provides an adequate spread of scores, and suggests that an algorithm that can capture both linear and nonlinear trends in the data may be more informative. Furthermore, there may be alternative descriptions of visual behaviour that may reflect the participant's cognitive biases and affective responses. Examples include statistical descriptions of oculomotor events (e.g., average fixation and saccade duration) across the entire page rather than towards each post (e.g., as used in Hoppe, Loetscher, Morey, & Bulling, 2018). As discussed in the methodology chapter, a main drawback of traditional psychological approaches is that fully interpretable, neurobiologically plausible models may not provide the best possible predictive power. As such, additional nonlinear associations between the metrics investigated and the outcome variable may exist that were not captured by this first study.

Thus, for the next study (Chapter 4), I decided to consider both linear and nonlinear trends by applying a selection of machine learning algorithms (detailed in the methodology chapter; see Section 2.4.4) to learn the association between a range of oculomotor metrics (see Section; 2.6) and the big five personality traits (Costa, 1996); a set of scores that are established within the literature as being predictive of behavioural outcomes (McCrae & Costa, 2004), and are expected to follow a normal distribution of scores (Rentfrow, Jokela, & Lamb, 2015). This builds on the previous study by making use of the same naturalistic (but constrained) type of web page stimuli while expanding the range of oculomotor metrics considered within a machine learning approach.

With this new ability to capture nonlinear trends and a different outcome variable, aspects of an individual's personality could be predicted with reasonable accuracy in certain cases (e.g., 74% for trait Openness, where chance was at 50%) from their eye movements while browsing the social media style web page. The findings of this study highlighted that machine learning techniques (which are not limited to neurobiologically plausible models) were able to predict the personality traits of Openness, Extroversion and Neuroticism significantly better than chance, although due to our implementation this was without granularity; only revealing if the participant is a high or low scorer upon the given trait. This finding builds upon well established literature finding that a wide range of personal attributes can be predicted from behaviour upon social networking sites (Azucar, Marengo, & Settanni, 2018; Kosinski, Stillwell, & Graepel, 2013; Settanni, Azucar, & Marengo, 2018), which social networking sites use to target users with advertisements. This is important as research suggests that targeting individuals with advertisements personalised to their personality is an effective strategy, with personality congruent advertisements leading to up to 50% improvement in purchases (Matz, Kosinski, Nave, & Stillwell, 2017). As such, this second study provides a proof-of-concept that commercially valuable personal information (Matz et al., 2017), may be unknowingly disclosed by individuals who enable their eye movements to be recorded while browsing social media style content.

Furthermore, the results from this study imply that it is not essential to know the content of the page to make reliable predictions about whether an individual is low or high upon each of the outlined personality traits. I discovered this by using the constrained nature of our artificially produced stimuli to compare content and location based descriptions of eye movement. From this analysis, I found that the classifiers trained upon statistical descriptions of eye movements (which can be calculated without knowing the page content) were able to predict a much wider range of personality traits (Openness, Extroversion and Neuroticism) significantly better than chance (and with greater accuracy) compared to classifiers trained upon either the content or location based metrics which require knowledge of the page content and layout. In comparison, none of the classifiers trained upon the content-based metrics performed significantly better than chance and a single personality trait (Openness, 69% Accuracy) was able to be predicted significantly better than chance by classifiers trained upon the location-based metrics. In regard to the main research question, this finding suggests that personal attributes may be predicted from the eye movements of users while they browse online social networking site (SNS) content even without the researcher knowing what they are viewing. This reduces the barrier to implement these techniques, although an important limitation is that this result may only be valid within situations where the participant's visual behavior is aggregated across repeated viewings of SNS-style stimuli.

The above limitation highlights that aspects of ecological validity were compromised to gain experimental control, and that this limits how far the results of this study can be generalisation to everyday browsing behaviour. This is an important point, as social media sites provide a novel visual environment with many features (e.g., scrolling and interactive elements) that were not present within the previous vision literature used to establish the link between visual behaviour and cognitive evaluations (i.e., preference decisions, see; Holmes et al., 1998; Nummenmaa et al., 2006; Shimojo et al., 2003). My mock social media web site contributed to this literature by capturing some of these elements (i.e., scrolling and content layout), but it failed to capture others. Content encountered upon social media sites is personalised based upon the user's interests, social circle, and previous actions which leads to a unique experience for each user. This crucial aspect was not replicated in my previous studies. Furthermore, the visual stimulus employed was not entirely ecologically valid, as compromises (i.e., non-scrolling side bars) were made to ensure a smooth interaction experience by eliminating the perceivable delay before the visual scene updated during scrolling events. In my third study (Chapter 5), I sought to apply my methodology within a truly ecologically valid stimulus; the participant's own Facebook News Feed.

From the results of my third study, I established that above chance (here chance is 33%) predictions for the personality traits of Extroversion (53% accuracy) and Conscientiousness (43% accuracy) can be achieved using eye movement data acquired from a single twenty second interaction with the participant's own Facebook News-Feed content. While displaying seemingly weaker findings than the first machine learning study (i.e., Chapter 4), in this study the classification task was much more challenging for two key reasons. Firstly, I pursued a three-category outcome to increase the granularity of information gained from the analysis (i.e., discovering if the participant was 'low', 'medium', or 'high' upon the given trait, rather than 'low' or 'high'). Secondly, the lack of experimental control, deliberately introduced in the pursuit of ecological validity, means that a substantial proportional of the individual differences in oculomotor behaviour may be attributable to variance in the low-level features of the visual scene (e.g., contrast, colour) rather than reflecting a gaze-trait association. As such, the algorithm must find stable patterns of oculomotor behaviour that generalise to new (i.e., out of sample) cases from amongst the variance due to external factors. Given these conditions, it is remarkable that the classifiers were able to perform significantly better than chance when provided only twenty seconds of data.

To gain further insight into which descriptions of eye movements gave insight into which personality traits, as in the previous study, I compared the performance of classifiers trained upon statistical descriptions of eye movements against classifiers trained upon content-based descriptions. Only trait Conscientiousness was predicted significantly better than chance when using the content-based descriptions (42.8% accuracy). Similarly, only trait Extroversion was predicted significantly better than chance when using statistical descriptions of eye movements (53.3% accuracy). As such, my third study builds upon the previous study to suggest that, within ecologically valid environments, content and statistical based descriptions of eye movement behaviour provide distinct insights into the participant's Conscientiousness and Extroversion personality category (low, medium, high) respectively. An important commonality between this study and the previous is the above-chance prediction of trait Extroversion, as when considered together, these two studies show that statistical descriptions of eye event parameters can be used to accurately predict an individual's trait extroversion category within both experimental and ecologically valid SNS visual environments. This confirms the earlier suggestion that trait Extroversion can be predicted from users browsing naturalistic SNS style visual environments, without the researcher knowing the content that is displayed upon the page.

An important dissimilarity is the reduced set of personality traits (Extroversion and Conscientiousness) that could be predicted significantly better than chance from visual behaviour compared to both my previous study, and previous literature within locomotive (Hoppe et al., 2018) and static image paradigms (Berkovsky et al., 2019).

This discrepancy may be linked to aspects of the experimental paradigm, as it is possible that Openness and Neuroticism manifest most strongly in the visual behaviour expressed within repeated presentation designs. However, given that a broad array of personality traits have been predicted from visual behaviour within locomotive conditions (Berkovsky et al., 2019) and static image paradigms (Berkovsky et al., 2019; Rauthmann, Seubert, Sachse, & Furtner, 2012), it is perhaps more likely that the results reflect either the increased difficulty of the classification task, or that the content of the visual scene is important. As such, when considered together, my findings suggest that social media coupled with machine learning is not a panacea for eliciting visual behaviour that is informative of aspects of the individual's personality, and that variation in context (i.e., NewFeed or a third party home page) may lead to better or worse predictive performance for different traits.

As discussed in Chapter 5 an alternative mechanistic interpretation is to conceptualise browsing Facebook as a information search task. Al-Samarraie, Eldenfria, and Dawoud (2016) found visual behaviour during information search tasks as being informative of Conscientiousness, Extroversion and Agreeableness; which provides a very similar pattern of results as found within chapter 5 with the exception of Agreeableness. Agreeableness is linked to how predisposed an individual is to acquiescence (Costa, 1996). As our participants were engaged in a freeviewing task, rather than a directed search task with a clear objective, there was no need for participants to comply with task directives. This may explain why the trait of Agreeableness was found by previous literature which provided a directed search paradigm (Al-Samarraie et al., 2016; Hoppe et al., 2018), but not replicated within Chapter 5. As such, I speculate that Agreeableness is most likely to influence the individual's visual behaviour when they are required to choose whether to accept a particular source of information during a search task. Due to this, it is perhaps not surprising that our results did not exactly replicate the findings of Al-Samarraie et al. (2016). This highlights the importance of conducting further research to empirically explore which categories of content, and contexts, are most informative of particular personal attributes. This will help inform the literature with an understanding of the risk posed to the user when sharing their eye movements within different contexts within and beyond SNS environments. For example, experimentally manipulating the browsing task to investigate if the inclusion of an acceptance criterion is essential for trait Agreeableness to be reflected in visual behaviour would inform the literature upon which types of context would pose a risk of the individual unknowingly disclosing information about this personality trait.

Finally, it appears that not only personality is reflected in an individual's eye movements whilst browsing SNS content; in my fourth and final study I identify that the additional personal attributes of narcissism and self-esteem category ('low', 'medium', 'high') along with the participant's sex ('female', 'male') can also be predicted from the oculomotor data collected as part of study three. Together, these findings (reported in Table 7.1) highlight that a broad array of personal attributes are able to be accurately predicted from eye movements while participants browse SNS content. This illustrates that additional unexplored personal attributes may be predictable from eye movement behaviour within this environment, and that the articles enclosed within this thesis are only a beginning of the exploration of this paradigm. It is worth noting that the number of observations collected within the third study represent only a fraction of that available to the providers of SNS content, and thus (since providing more observations to learn from usually improves the performance of machine learning models) likely underestimates what is possible from the paradigm.

TABLE 7.1: Summary of Significant Results from the Own Facebook Data Set (Chapter 5 and 6)

| Personal Attribute | Category (Number, Strategy) | Accuracy | F1 / F1 Macro |
|---|---|---|---|
| Conscientiousness | N=3, Quantile | 42.80% | 0.398 |
| Extroversion | N=3, Quantile | 53.30% | 0.519 |
| Narcissism | N=3, Quantile | 47.80% | 0.425 |
| Self-Esteem | N=3, k-means | 57.20% | 0.436 |
| Sex | N=2, Self-Assigned | 63% | 0.643 |

### 7.1.1 Implications

Together, our results indicate that the substantial increase in the velocity and volume of data that accompanies eye tracking techniques (compared to the digital footprints traditionally investigated, as mentioned in the introduction) allows aspects of a user's personal attributes to be predicted better than chance from a single twenty second interaction. This result was achieved by aggregating data over substantially less time than reported in previous literature wishing to predict personality trait scores from users online behaviour (e.g., likes on Facebook; Kosinski et al., 2013), where the information captured often reflects weeks, months or years of social media use. As highlighted in previous literature (Kosinski, Matz, Gosling, Popov, & Stillwell, 2015), the ability to predict personal attributes from visual behaviour has important implications for privacy. Large entertainment companies such as Facebook, Netflix, and YouTube have invested in redesigned interfaces that allow for eye based interaction (Coldewey, 2020), with two of the largest companies, Google and Facebook, purchasing eye tracking companies (Dickson, 2017). Eye tracking has also been increasingly integrated into virtual reality (VR) headsets both for academic research (Clay, König, & König, 2019) and consumer use, with increasing commercial interest in techniques such as foveated rendering (Patney et al., 2016), which is reliant on in-built eye tracking. As such, there is a growing potential for the layperson to encounter and engage with eye tracking technologies in everyday life. However, it is still currently unreasonable to expect consumers to understand that personal information is decodable from visual behaviour. As such, by providing their visual behaviour to a service provider, consumers may be unknowingly disclosing information about aspects of their personality and other psychological attributes.

This provides new opportunities for invasion of privacy, and I caution that the depth of information available from eye movements is still currently unknown, but our work has hinted at the possibilities. Of particular interest is that our results show that it is not essential to know what is being displayed upon the screen for above-chance predictions of an individual's, sex, and psychological factors (i.e., whether they are low, medium or high upon the traits of narcissism, self-esteem, extroversion and conscientiousness). Without the correct legislative framework, this ability for service providers to collect, analyse, and infer attributes from individual's without their knowledge or consent is deeply troubling. However, while it is clear that eye movement data provides a novel and rather immediate method of invading an individual's privacy (i.e., making assumptions about the individual without their knowledge or explicit consent), it may be worth considering if this is a lesser evil.

To elaborate, predicting an individual's attributes from statistical descriptions of eye movements (i.e., that are not content-bound) does not require that the researcher undertake a detailed analysis of months or years of the user's digital footprint, as is required by current techniques (e.g., as in Azucar et al., 2018; Eftekhar, Fullwood, & Morris, 2014; Gosling, Augustine, Vazire, Holtzman, & Gaddis, 2011; Kosinski et al., 2013). Thus, the privacy of the user is relatively preserved. The counterpoint to this is that a user can choose to interact with a social media site (e.g., to click 'like' or share a piece of content), but they may find it harder to control their eye movements, as eye movements are often involuntary (Hayhoe, Mennie, Gorgos, Semrau, & Sullivan, 2004). Thus, a direction for future research is to develop an understanding of whether individuals would object less to their personal attributes being predicted from their eye movements, compared to predictions made from their previous browsing history.

Finally, there are distinct advantages that emerge from the trend towards integrating eye tracking information into human computer interactions. One key aspect is to provide the opportunity for novel human computer interactions that enhance accessibility options, which may promote a wider audience to engage with online services (Menges, Kumar, & Staab, 2019). Furthermore, VR environments enabled with eye tracking technology provide a new frontier of opportunities for enabling early clinical interventions (Hemmingsson, Ahlsten, Wandin, Rytterström, & Borgestig, 2018). For example, when combined with machine learning techniques, eye movements in VR may provide an opportunity for widespread screening for conditions such as Parkinsons, where early detection is critical for establishing better quality of life outcomes. Thus, privacy concerns should not necessarily mean ruling out all uses of eye tracking combined with social media (Tseng et al., 2013).

## 7.2 Contributions to Methodological Practice: Data Driven Category Formation

A wide range of previous literature has sought to predict personal attributes from oculomotor behaviour, and transformed the continuous questionnaire scores acquired from established psychological scales into categorical outcomes to do so (e.g., Berkovsky et al., 2019; Hoppe et al., 2018; Settanni et al., 2018). However, rarely is a justification given for the number of categories formed, nor the method used to form them. Just as in hypothesis testing, where multiple tests must be accounted for, the potential for authors to repeatedly attempt to predict the output from different categorisation techniques until a significant association is found is particularly worrisome. This follows widespread concerns over the reproducibility of machine learning results (Tatman, VanderPlas, & Dane, 2018), which is a similar dilemma as the psychological reproducibility crisis (Aarts et al., 2015).

A simple solution is to include the formation of categorical outcomes in the machine learning pipeline as a preprocessing stage, whereby the optimal strategy can be chosen (via cross-validation) that maximises the performance of the machine learning algorithm. This provides transparency and a way to identify optimally performing categorisation strategies. However, this is likely to substantially increase the computational cost of the analysis and does not guarantee that the categories will be semantically meaningful or optimal for the application. For example, in a situation where the researcher wishes to gain knowledge about the individual's attributes from a set of features, the researcher may be most interested in a certain granularity of the predicted information (e.g., having more fine grained categories). As increasing the number of categories makes the classification task more challenging (and thus likely to lead to reduced performance), high granularity would not likely be selected by optimising the categorisation strategy within a cross-validated preprocessing stage. However, a researcher viewing this data may be willing to sacrifice accuracy to achieve additional granularity (i.e., having a greater number of categories), and thus opt for choosing a non-optimal strategy performance wise. Alternatively, the researcher may have domain specific knowledge that they can apply to the problem, to ensure that the categories formed are meaningful to the given application.

To empower the researcher to choose their preferred solution, in this thesis (Chapter 6) I have also explored how to approach the formation of categorical outcomes from continuous questionnaire scores in a principled manner. Importantly, I highlight the importance of considering the validity (i.e., appropriateness) of the categories formed in regard to the research question and suggest that alternative techniques to quantile based methods should be applied to assign continuous personality trait scores to categorical outcomes only upon three conditions; that the silhouette score of the categories is meaningfully improved upon, that enough observations are present in each category for the algorithm to learn from, and that the categories retain content validity (i.e., reflect groupings that are meaningful in the context of the analysis). As such, our methodology provides a way for the researcher to, in a data-guided manner, justify the choices they have made. This promotes transparency in the decision making process underlying the choice of, for example, two or three category outcomes, while allowing for the researcher to utilise their domain knowledge and expertise to preserve content validity.

A potential critique of this strategy is that the researcher can not know *apriori* whether the chosen solution will result in strong classifier performance. Therefore, as an alternative strategy that seeks to balance these two interests (validity and performance), a useful recommendation would be a hybrid approach, whereby the researcher includes the categorisation strategy as a preprocessing stage to highlight the top performing strategies, and then conducts an investigation (applying the visualisation techniques and methods described within Chapter 6) to ensure that the resulting structure is meaningful for the application at hand. This seeks the best performing solution that is meaningful, and prevents the undesirable situation whereby the results of the classification look excellent upon the performance metric, but give little insight into behavior (e.g., by constructing categories that are not meaningful). However, I note that the above process selects the first meaningful solution that performs well, which may not be the most meaningful solution (i.e., the researcher is seeking to find a middle ground). This illustrates how the work presented can be extended to accommodate researchers who seek to prioritise performance, which is different to the emphasis upon finding coherent categories as pursued within the original paper.

## 7.3 Limitations and Future Directions

Equally important to raising awareness about the potential for disclosing personal information via eye tracking techniques is to avoid overstating what is currently possible. In the most ecologically valid study presented as part of this thesis, I collected eye movements from a modest cohort of 180 participants while they browsed their own Facebook NewsFeed content, and found that a variety of personal attributes could be predicted significantly better than chance from a single twenty second interaction (outlined in Table 7.1). It is important to note that the predictive performance of the models is modest in comparison to literature that seeks to predict personal attributes from digital recordings of SNS use. For example, Kosinski et al. (2013) reports 93% accuracy for classifying whether the user was male or female based upon their profile of Facebook 'likes', which substantially outperforms our reported 63% accuracy. Furthermore, in their meta-analysis, Azucar et al. (2018) identifies that each of the big five personality traits can be predicted significantly better than chance from various digital traces, which illustrates that a wider set of outcomes can be predicted from currently exploited digital footprints than what is known to be possible from eye movements alone.

The results presented within this thesis show that eye movements alone provide only a broad insight into (e.g., whether someone is low, medium, or high) the personal attributes investigated. This pales in comparison to the resolution afforded by traditional psychological questionnaires (Costa, 1996), and alternative digital footprints (e.g., Kosinski et al., 2013). As such, I can conclude that the main risk to users face when providing their eye movements is not related to the accuracy or breadth of information that can be predicted from this data. Instead, it relates to the unprecedented ability for a researcher to gain a general (above-chance) insight into a subset of personal factors from a short single interaction.

Finally, while eye movements alone might not be highly informative of personal information, Azucar et al. (2018) highlights that combining multiple sources of digital footprints often leads to better classifier performance when predicting personality traits. As such, a key direction for future research is to explore whether combining eye movements with existing digital records allows additional insight (i.e., granularity and precision) when predicting personal attributes from interactions with SNS content.

When compared to previous literature, the findings of Chapter 5 and 6 (Table 7.1) are comparable to those of Hoppe et al. (2018), who also employed machine learning approaches to predict the personality traits (categorised as low, medium or high) of participants from their eye movement behaviour in a naturalistic setting. The authors found that extroversion (48.6%) and conscientiousness (43.1%) could be predicted significantly above chance. Importantly, Hoppe et al. (2018) recorded the participant's visual behaviour across a changing visual scene (i.e., participants did not all view an identical scene), which makes the difficulty of the classification task similar to the literature presented within this thesis, which may explain the comparable findings. I note, however, that Hoppe et al. (2018) additionally found neuroticism (40.3%) and agreeableness (45.9%) to be predicted significantly better than chance, which was not the case in my studies. As such, it is perhaps fair to state that my research has illustrated that certain facets of personality, rather than personality as a whole, can be predicted from eye movements upon SNS content.

Two key differences, and thus likely candidates for the difference in findings between my studies and Hoppe et al. (2018), relate to the amount of time the participant's eye movements were recorded for and the type of metrics used to describe the eye movements. In Hoppe et al. (2018), the participants eye movements were recorded for ten minutes, which is a much longer time period than the twenty seconds investigated in our paradigm. As such, it is possible that the traits of neuroticism and agreeableness are best expressed in visual behaviour captured over an extended duration of time, and thus were not captured by the twenty second study on account of this. A direction for future research would be to reproduce our study over a ten minute viewing duration, and conduct the analysis over different periods of time. An additional advantage of such research would be to provide data on the relative insight provided for each trait as the participants viewing duration increased, allowing an understanding of the minimum duration required for a given accuracy of prediction (for each personality trait, or a given personal attribute).

Alternatively, it may be that the visual behaviour informative of trait agreeableness and neuroticism is captured by the metrics utilised in Hoppe et al. (2018). This is as the features used by Hoppe et al. (2018) differ both quantitatively (i.e., 207 features, vs $\leq$ 28 features for my studies) and qualitatively (e.g., the use of n-gram features) from those employed in my studies. Notably, the most informative metrics reported for both agreeableness and neuroticism were n-gram features (detailed within the introduction chapter), which were not utilised within my studies.

Thus, a direction for future research is to employ a feature-group style approach to distinguish if the time element, alternative features, or both, lead to the differences in findings between my studies and Hoppe et al. (2018). If neither time or metric differences can account for our different findings for these two personality traits, then it may be that the differences in stimuli (i.e., locomotion/shop environment in Hoppe et al., 2018) are responsible for the differences. Namely, although personality trait-congruent behaviour is known to be expressed upon social media sites (Azucar et al., 2018), it may be expressed most strongly when walking around and during face-to-face interactions with other people.

When comparing our findings to results from literature conducted within more controlled lab conditions with specially selected visual stimuli and tasks, the relevant studies come from Bargary et al. (2017) and Berkovsky et al. (2019). Bargary et al. (2017) investigated a large cohort of 1058 participants while they completed four visual tracking tasks, and found a weak linear association between trait Extroversion and the participant's eye movements. Bargary et al. (2017) also found significant differences between men and women for several oculomotor metrics (e.g., during smooth pursuit), but did not investigate whether a classifier could use this information to make accurate predictions. As such, I can not directly compare the sex prediction results to this literature. However, when compared to the personality trait results reported in this thesis (reported in Table 7.1), it is clear that my results exhibit greater predictive power than those of Bargary et al. (2017), and successfully predict a wider range of personality traits. Importantly, the stimuli shown by Bargary et al. (2017) lacked affective or social information, which, as discussed in the introduction, may have limited the ability of the participants eye movements to reflect their learnt associations with objects within the scene. This, together with Bargary et al. (2017) utilising linear methods, which (as shown in study one) may underestimate the predictive utility of eye movements, may provide an explanation for the difference in findings.

While outperforming Bargary et al. (2017), our prediction performance results are much more modest when compared to Berkovsky et al. (2019), who reports accuracy over 80% for each of the big five personality traits within a three category (low, medium, and high) classification task. This increased performance may be attributable to the authors showing a series of emotive images within a static image presentation paradigm, which as discussed reduces the difficulty of the classification problem. However, due to the small sample size along with additional methodological issues (explored in detail in the introduction and methods), I am cautious about drawing conclusions from this literature. An important direction for future research is to replicate the Berkovsky et al. (2019) study within a larger cohort, to ensure that the results generalise.

Finally, a barrier that limits our insight into the risk of unknowing disclosing personal information through sharing eye movements upon SNS content is the disparity between the quality of the eye tracking equipment used in my experiments, and the type of eye tracking that is likely available for use in everyday situations. As devices with lower sampling rates are most likely to be encountered by individuals' browsing social media from home, the use of laboratory grade eye tracking equipment in my studies remains an obstacle to generalising our results directly to everyday users. Given that it is currently unknown whether there is any risk to the user when being tracked using lower-resolution hardware, an essential direction for future research is to provide empirical evidence to explore whether the methods used within this thesis generalise to low-resolution settings. This is particularly important in light of recent literature, which suggests that a deep-learning approach may be able to track the eye movements of users (with accuracy rivalling that of industry-grade head mounted equipment) using just a mobile phone's 'selfie' camera (Valliappan et al., 2020).

The inherent disadvantage of lower-resolution (e.g., 50-100hz vs 300-1000hz) eye tracking hardware is that a larger temporal gap between samples means less precision when measuring the onset and end (and thus duration) of eye movement events. However, as we can assume a uniform distribution for eye movement events (i.e., the event is equally likely to fall anywhere within the sampling time-frame; Andersson, Nystrom, & Holmqvist, 2010), central limit theory dictates that as the number of captured eye movement events rises, the statistical descriptions based upon these events will approximate the accuracy provided by higher-sampling devices. This is important because statistical descriptions of eye movements (based upon averaging across numerous events) have been used successfully within this thesis to predict a wide range of personal information; as such, it is reasonable to assume that statistical descriptions of visual behaviour using lower-resolution eye tracking devices will be informative of personality, but may require a longer duration of recording to achieve the same accuracy as higher-resolution systems. This provides a theoretical basis for results based upon statistical descriptions of eye movement events generalising within the context of lower-resolution eye tracking devices.

## 7.4 Conclusion

In this thesis I found that when viewing social media web pages, participants' visual behaviour reflects an efficient search biased towards both the center of the page (where the 'posts' are) and the faces contained within the stimuli. This suggests that describing how the participant visually interacts with each 'post' on the page is likely to capture the majority of the visual behaviour upon the page, and provides some validity to my approach in defining regions of interest across my empirical studies. This allowed me to address the hypothesis that personal attributes may be reflected in an individual's eye movements whilst they browse social media, as emphasised in the first chapter (which details how individual preferences can subconsciously bias the expression of visual behaviour). The empirical studies contained within this thesis provide supporting evidence for this hypothesis by showing that a wide range of personal attributes can indeed be predicted from an individual's visual behaviour, and a key contribution of this thesis is to highlight that personal information can be predicted at above-chance accuracy within very short time scales due to the increased velocity and volume of data afforded by eye tracking techniques. This is unique from previous approaches based upon digital footprints, which require information agglomerated over much longer time periods.

Whilst the quality of the predictions varies across studies, in general the machine learning approach can be seen to perform well even when the linear regression methods predict little variance. For example, in Chapter 3 I found that, despite the eye movement behaviour of the participants being focused upon the central posts, traditional linear techniques explained very little variance (5-10%) in the outcome variable of how much the participant enjoyed viewing the presented posts. Yet the same data set using machine learning techniques yielded reasonably accurate predictions of the individual's personality in Chapter 4. Thus, a key theoretical implications is that future research should consider employing machine learning methodologies when they are interested in out-of-sample performance. In particular, conducting a wide-scale empirical investigation using machine learning techniques upon the eye movements of participants browsing their own Facebook Newsfeed (chapter 5), allowed me to identify that, within an ecologically valid stimulus, it is mainly Conscientiousness and Extroversion that can be predicted from eye movement behaviour. The differences in findings between Chapter 4 and 5 (i.e., Openness and Neuroticism were not predicted above chance in Chapter 5, but were in Chapter 4) are thought to arise due to the methodological differences between a controlled experimental stimulus, and an ecologically valid social media environment. As such, a key theoretical contribution of this thesis is that a key driver of which personal attributes are reflected in eye movements is the type of stimulus being presented to the individual. Furthermore, in addition to depending on the stimulus shown, the category formation technique can lead to varying classifier performance. Future literature wishing to investigate the potential risk of unknowingly disclosing personal information via eye tracking technology must consider the ecological validity of the visual stimulus to be of paramount importance, and provide a rationale when using a given category formation technique; without this it will be difficult to develop a consistent and comprehensive body of research upon this topic. This builds upon the comments of Tatler, Hayhoe, Land, and Ballard (2011) who have previously mentioned that theory and models built upon simplified stimulus sets may not generalise outside of the experimental context, and provide misleading predictions for naturalistic stimuli. It is also important to highlight that social media content is not a panacea for eliciting trait-congruent visual behaviour, as only a subset of the investigated personal attributes were predictable.

A shortcoming in the proposed theory is that it is unknown whether it important to have a directed search paradigm to elicit trait Agreeableness congruent visual behaviour (e.g., as suggested in Chapter 5), and whether information seeking paradigms reliably elicit trait Conscientiousness and Extroversion congruent visual behaviour (e.g., as suggested by comparing the results of this thesis to; Al-Samarraie et al., 2016; Hoppe et al., 2018). This leads me to suggest that a key future direction for research is to investigate under which contexts a given trait is reflected in the participant's visual behaviour, as this provides additional insight into which social media content (or situations) pose the greatest risk to the participant's privacy.

In this thesis I have also highlighted that the opportunities for participants to share their eye movements are likely to become more frequent; with companies such as Facebook investing in eye tracking companies and eye based interfaces (Coldewey, 2020), and the emergence of new mobile-based eye tracking methods (Valliappan et al., 2020). Furthermore, I have demonstrated that it is not essential for the researcher to have knowledge of what the participant is viewing upon the social media page to make predictions about aspects of the individual's personality (Conscientiousness, Extroversion, Narcissism, Self-Esteem) and sex. This illustrates that even if the individual does not choose to share what is upon their screen, a wide range of assumptions can be made about them without their explicit knowledge, which may allow for the user to be targeted with persuasive material based upon their assumed psychological profile (Matz et al., 2017). As such, this thesis has uncovered that substantial invasions of privacy are possible using eye based methods. However, future research is required to understand how the public reacts to this knowledge as, in some situations, the techniques employed within this thesis may be a way of preserving an individual's privacy (i.e., allowing inferences to be made without the disclosure of weeks, months, or years of digital footprint).

Finally, before the results of this thesis can be confidently applied to real world social media use cases, it is imperative that the findings are replicated using eye tracking technology that users are likely to come into contact with. This is a constantly changing field, with recent rapid developments requiring that researchers keep up to date with the newest technology to accurately report the potential risk to users from sharing their eye movements.

## References


Aarts, A., Anderson, J., Anderson, C., Attridge, P., Attwood, A., Axt, J., … Zuni, K. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), aac4716. doi:10.1126/science.aac4716

Al-Samarraie, H., Eldenfria, A., & Dawoud, H. (2016). The impact of personality traits on users' information-seeking behavior. *Inf. Process. Manag. 53*(1), –. doi:http://dx.doi.org/10.1016/j.ipm.2016.08.004

Andersson, R., Nystrom, M., & Holmqvist, K. (2010). Sampling frequency and eye-tracking measures: How speed affects durations, latencies, and more. *Journal of Eye Movement Research*, *3*(3). doi:10.16910/jemr.3.3.6

Azucar, D., Marengo, D., & Settanni, M. (2018). Predicting the Big 5 personality traits from digital footprints on social media: A meta-analysis. *Pers. Individ. Dif. 124*, 150–159. doi:10.1016/j.paid.2017.12.018

Bargary, G., Bosten, J. M., Goodbourn, P. T., Lawrance-Owen, A. J., Hogg, R. E., & Mollon, J. (2017). Individual differences in human eye movements: An oculomotor signature? *Vision Res. 141*, 157–169. doi:10.1016/j.visres.2017.03.001

Berkovsky, S., Taib, R., Koprinska, I., Wang, E., Zeng, Y., Li, J., & Kleitman, S. (2019). Detecting Personality Traits Using Eye-Tracking Data. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (pp. 1–12). CHI '19. doi:10.1145/3290605.3300451

Clay, V., König, P., & König, S. U. (2019). Eye tracking in virtual reality. *Journal of Eye Movement Research*, *12*(1). doi:10.16910/jemr.12.1.3

Coldewey, D. (2020). Facebook, YouTube, Netflix and more get eye-tracking apps from Tobii.

Costa, P. T. (1996). Work and personality: Use of the NEO-PI-R in industrial/organisational psychology. *Appl. Psychol. 45*(3), 225–241. doi:10.1111/j.1464-0597.1996.tb00766.x

Dickson, B. (2017). Unlocking the potential of eye tracking technology. *TechCrunch*.

Eftekhar, A., Fullwood, C., & Morris, N. (2014). Capturing personality from Facebook photos and photo-related activities: How much exposure do you need? *Comput. Human Behav. 37*, 162–170. doi:10.1016/j.chb.2014.04.048

Gosling, S. D., Augustine, A. A., Vazire, S., Holtzman, N., & Gaddis, S. (2011). Manifestations of Personality in Online Social Networks: Self-Reported Facebook-Related Behaviors and Observable Profile Information. *Cyberpsychology, Behav. Soc. Netw. 14*(9), 483–488. doi:10.1089/cyber.2010.0087

Hayhoe, M, Mennie, N, Gorgos, K, Semrau, J, & Sullivan, B. (2004). The Role of Internal Models and Prediction in Catching Balls. *J. Vis. 4*(8), 156–156. doi:10.1167/4.8.156

Hemmingsson, H., Ahlsten, G., Wandin, H., Rytterström, P., & Borgestig, M. (2018). Eye-Gaze Control Technology as Early Intervention for a Non-Verbal Young Child with High Spinal Cord Injury: A Case Report. *Technologies*, *6*(1), 12. doi:10.3390/technologies6010012

Holmes, C. J., Hoge, R, Collins, L, Woods, R, Toga, A. W., & Evans, A. C. (1998). Enhancement of MR images using registration for signal averaging. *J. Comput. Assist. Tomogr. 22*(2), 324–333. doi:10.1097/00004728-199803000-00032

Holmes, T., & Zanker, J. M. (2012). Using an Oculomotor Signature as an Indicator of Aesthetic Preference. *i-Perception*, *3*(7), 426–439. doi:10.1068/i0448aap

Hoppe, S., Loetscher, T., Morey, S. A., & Bulling, A. (2018). Eye Movements During Everyday Behavior Predict Personality Traits. *Front. Hum. Neurosci. 12*, 105. doi:10.3389/FNHUM.2018.00105

Kosinski, M., Stillwell, D., & Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. *Proc. Natl. Acad. Sci. 110*(15), 5802–5805. doi:10.1073/pnas.1218772110

Kosinski, M., Matz, S. C., Gosling, S. D., Popov, V., & Stillwell, D. (2015). Facebook as a research tool for the social sciences: Opportunities, challenges, ethical considerations, and practical guidelines. *Am. Psychol. 70*(6), 543–556. doi:10.1037/a0039210

Matz, S. C., Kosinski, M., Nave, G., & Stillwell, D. J. (2017). Psychological targeting as an effective approach to digital mass persuasion. *Proceedings of the National Academy of Sciences*, *114*(48), 12714–12719. doi:10.1073/pnas.1710966114. eprint: https://www.pnas.org/content/114/48/12714.full.pdf

Maughan, L., Gutnikov, S., & Stevens, R. (2007). Like more, look more. Look more, like more: The evidence from eye-tracking. *J. Brand Manag. 14*(4), 335–342. doi:10.1057/palgrave.bm.2550074

McCrae, R., & Costa, P. (2004). A contemplated revision of the NEO Five-Factor Inventory. *Pers. Individ. Dif. 36*(3), 587–596. doi:10.1016/S0191-8869(03)00118-1

Menges, R., Kumar, C., & Staab, S. (2019). Improving User Experience of Eye Tracking-Based Interaction: Introspecting and Adapting Interfaces. *ACM Transactions on Computer-Human Interaction*, *26*(6), 37:1–37:46. doi:10.1145/3338844

Nummenmaa, L., Hyönä, J., & Calvo, M. G. (2006). Eye movement assessment of selective attentional capture by emotional pictures. *Emotion*, *6*(2), 257–268. doi:10.1037/1528-3542.6.2.257

Patney, A., Kim, J., Salvi, M., Kaplanyan, A., Wyman, C., Benty, N., … Luebke, D. (2016). Perceptually-based foveated virtual reality. In *ACM SIGGRAPH 2016 Emerging Technologies* (pp. 1–2). SIGGRAPH '16. doi:10.1145/2929464.2929472

Rauthmann, J. F., Seubert, C. T., Sachse, P., & Furtner, M. R. (2012). Eyes as windows to the soul: Gazing behavior is related to personality. *J. Res. Pers. 46*(2), 147–156. doi:10.1016/j.jrp.2011.12.010

Rentfrow, P. J., Jokela, M., & Lamb, M. E. (2015). Regional Personality Differences in Great Britain. *PLOS ONE*, *10*(3), e0122245. doi:10.1371/journal.pone.0122245

Settanni, M., Azucar, D., & Marengo, D. (2018). Predicting Individual Characteristics from Digital Traces on Social Media: A Meta-Analysis. *Cyberpsychology, Behav. Soc. Netw. 21*(4), 217–228. doi:10.1089/cyber.2017.0384

Shimojo, S., Simion, C., Shimojo, E., & Scheier, C. (2003). Gaze bias both reflects and influences preference. *Nat. Neurosci. 6*(12), 1317–1322. doi:10.1038/nn1150

Tatler, B., Hayhoe, M., Land, M., & Ballard, D. (2011). Eye guidance in natural vision: Reinterpreting salience. *J. Vis. 11*(5), 5–5. doi:10.1167/11.5.5

Tatman, R., VanderPlas, J., & Dane, S. (2018). A Practical Taxonomy of Reproducibility for Machine Learning Research. In *Reproducibility in Machine Learning*. ICML, Stockholm, Sweden. Retrieved September 15, 2020, from https://openreview.net/forum?id=B1eYYK5QgX

Tseng, P.-H., Cameron, I. G. M., Pari, G., Reynolds, J. N., Munoz, D. P., & Itti, L. (2013). High-throughput classification of clinical populations from natural viewing eye movements. *J. Neurol. 260*(1), 275–284. doi:10.1007/s00415-012-6631-2

Valliappan, N., Dai, N., Steinberg, E., He, J., Rogers, K., Ramachandran, V., . . . Navalpakkam, V. (2020). Accelerating eye movement research via accurate and affordable smartphone eye tracking. *Nature Communications*, *11*, 4553. doi:10.1038/s41467-020-18360-5

## Appendix A

# Supplementary Materials: Chapter Three

Here I provide the five versions of the animal themed webpages created for the study 'Does emotional valence predict oculomotor behaviour whilst browsing social media?' in chapter 3.

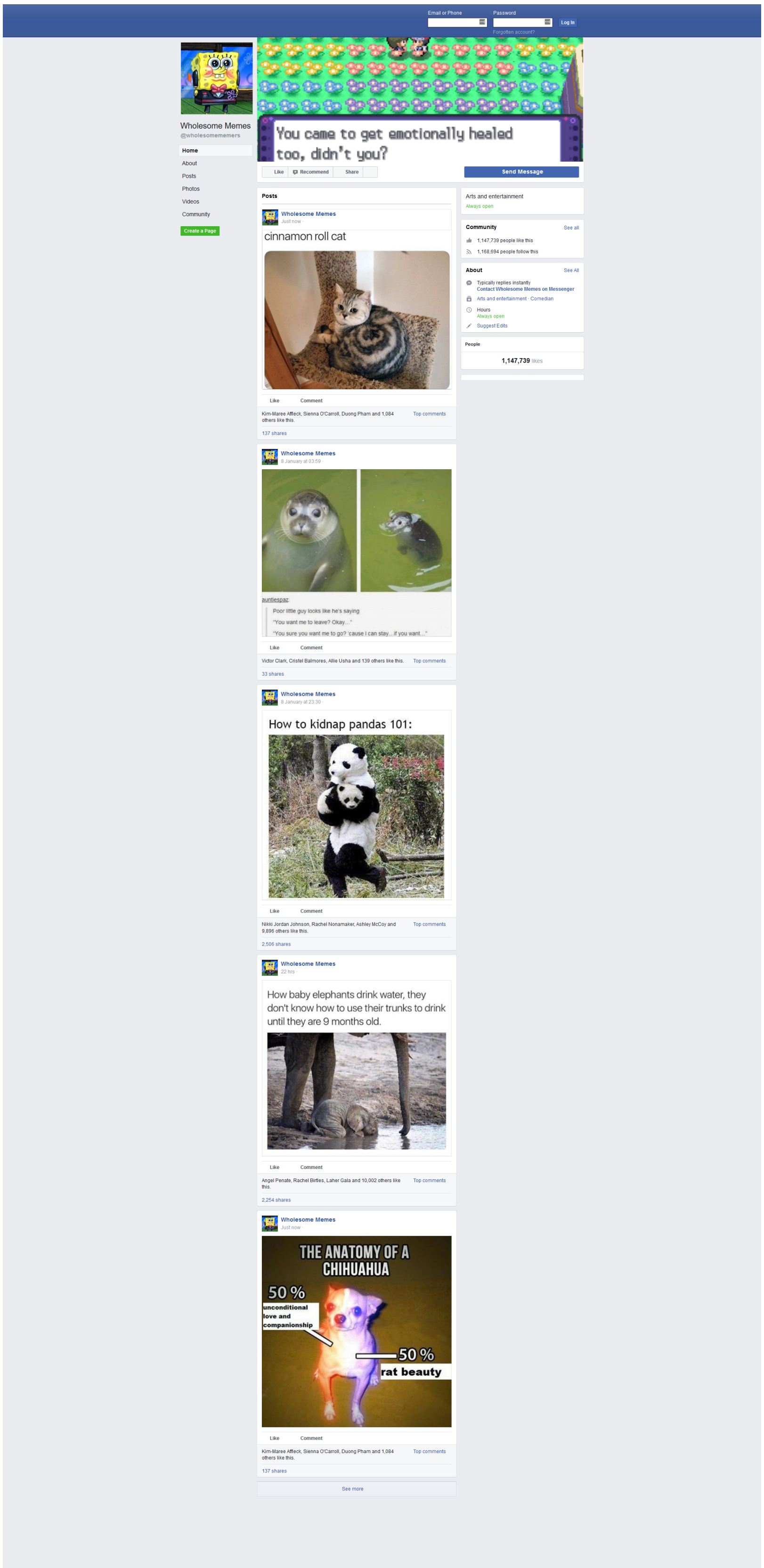


FIGURE A.1: Stimuli A: Animal composite webpage

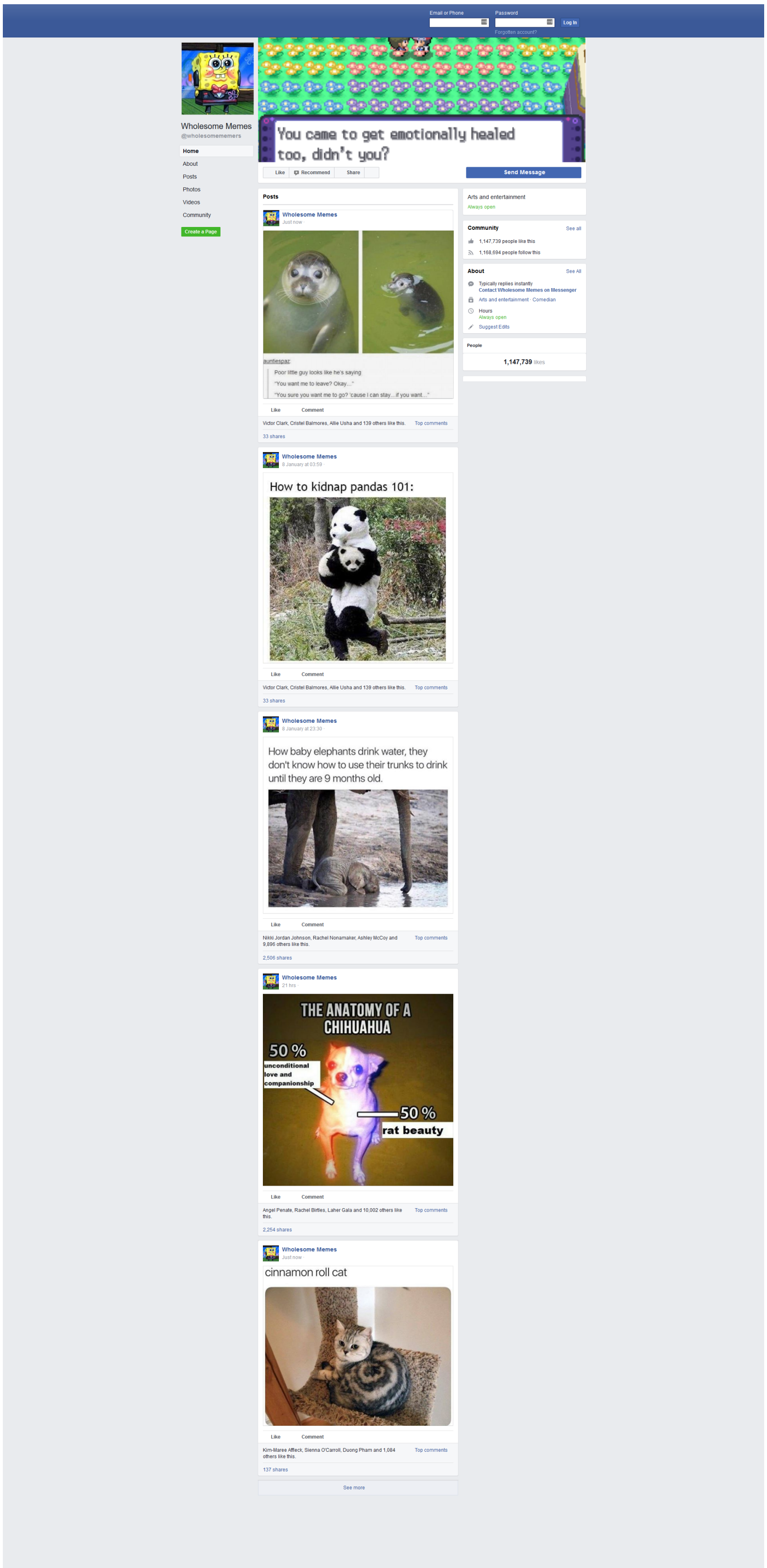


FIGURE A.2: Stimuli B: Animal composite webpage

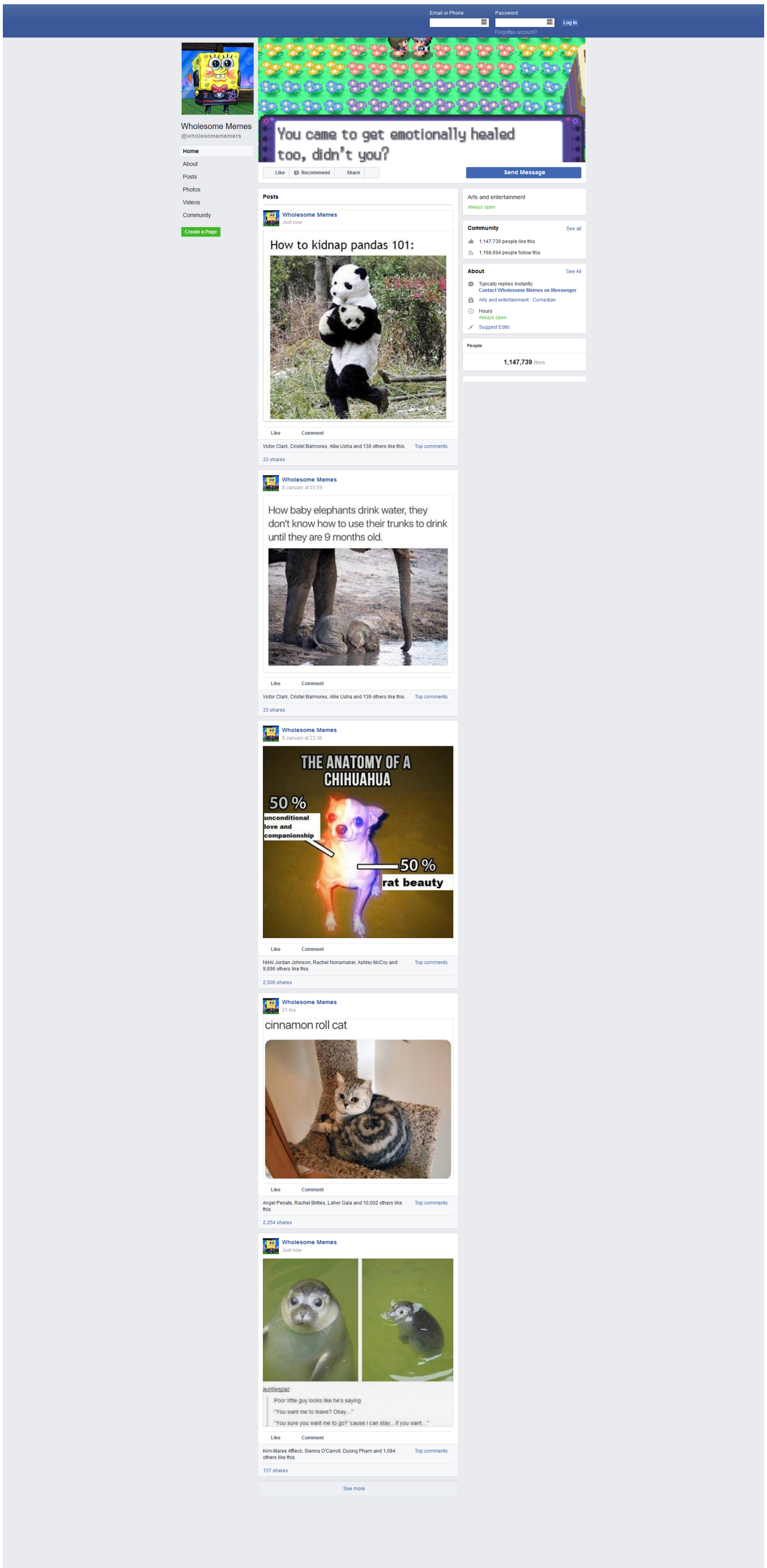


FIGURE A.3: Stimuli C: Animal composite webpage three

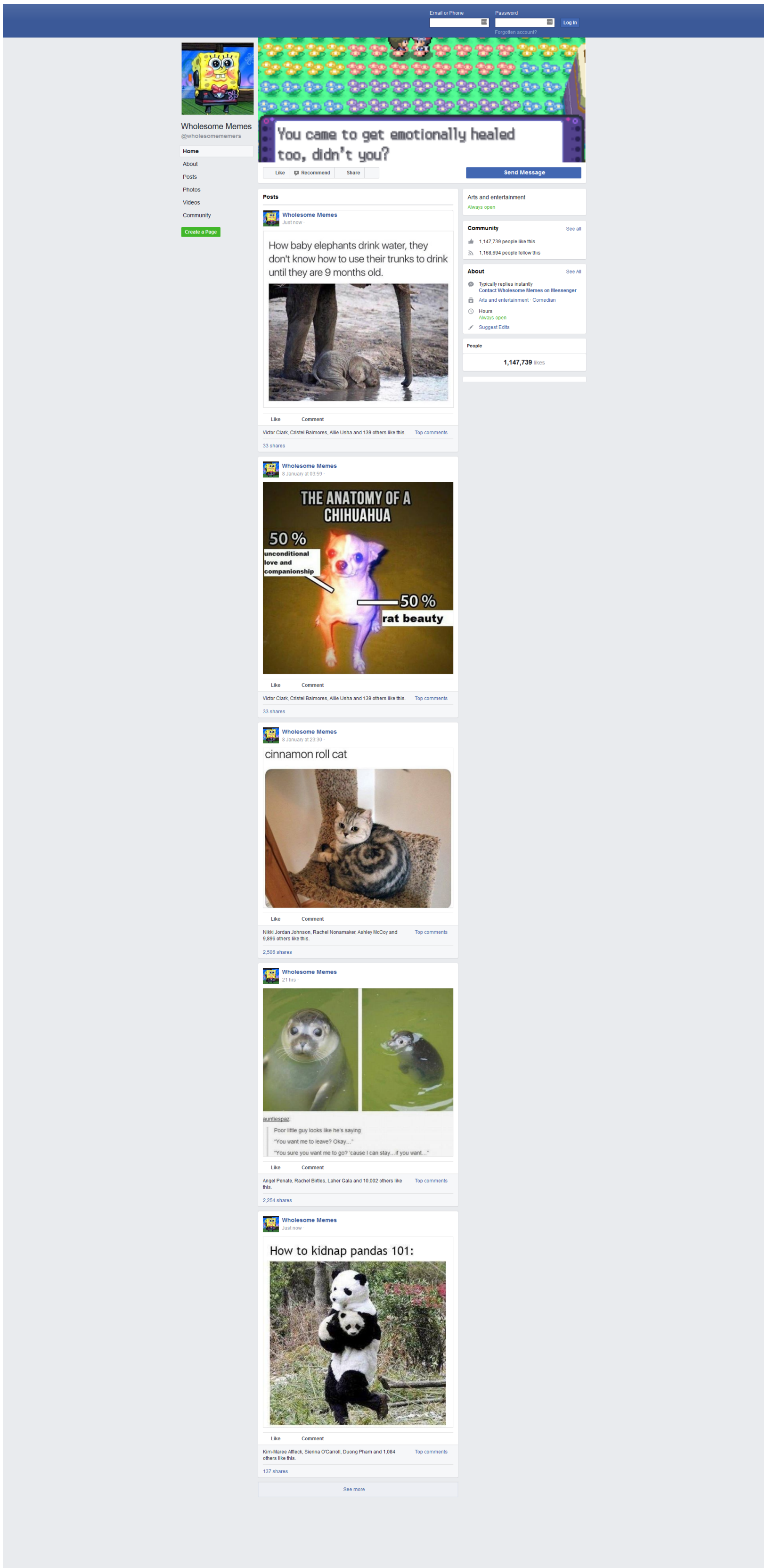


FIGURE A.4: Stimuli D: Animal composite webpage

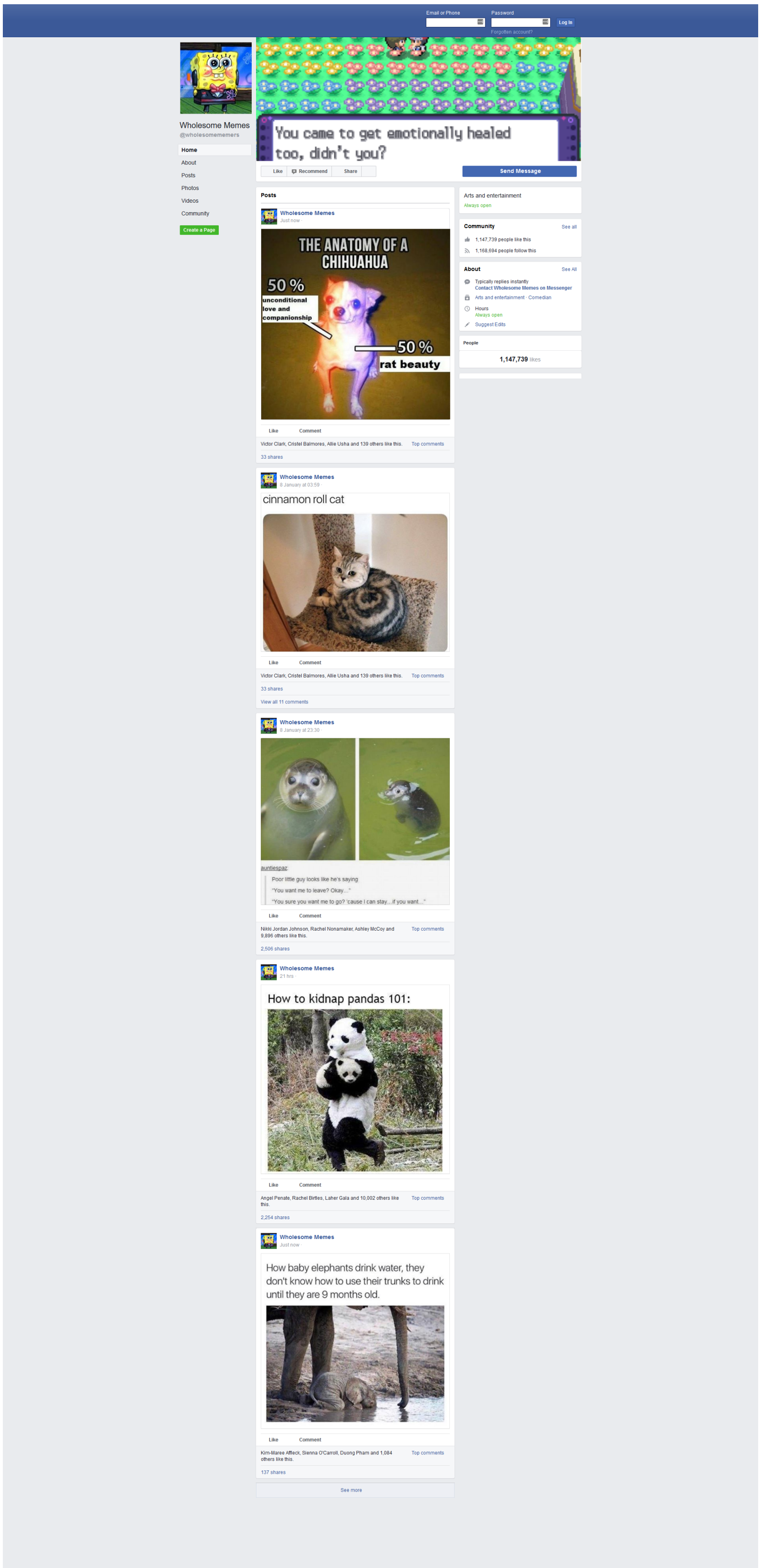


FIGURE A.5: Stimuli E: Animal composite webpage

# Appendix B

# Supplementary Materials: Chapter Six

## B.1 Study Two: Dendrograms for Trait Personality Scores

This supplementary materials section contains the dendrogram visualisations (based upon the agglomerative clustering technique) for the remaining big five personality traits of Openness, Extroversion, Agreeableness and Neuroticism.

## B.2 Study Three: Dendrogram and Silhouette Scores for Political Inclination, Self-Esteem and Narcissism

In the third study in chapter six, I investigated whether additional personal attributes (outside of the big five personality traits) could be predicted from eye movement behaviour. In order to form categorical outcomes, the procedure outlined in study two was followed. The following graphs highlight the results of first performing agglomerative clustering, then evaluating if applying k-means clustering using the suggested number of clusters leads to categories that are substantially more internally cohesive, and externally distinct (i.e., with greater silhouette scores). For the k-means approach to be employed over the quantile split approach, I required a non-overlapping advantage in silhouette score.

—

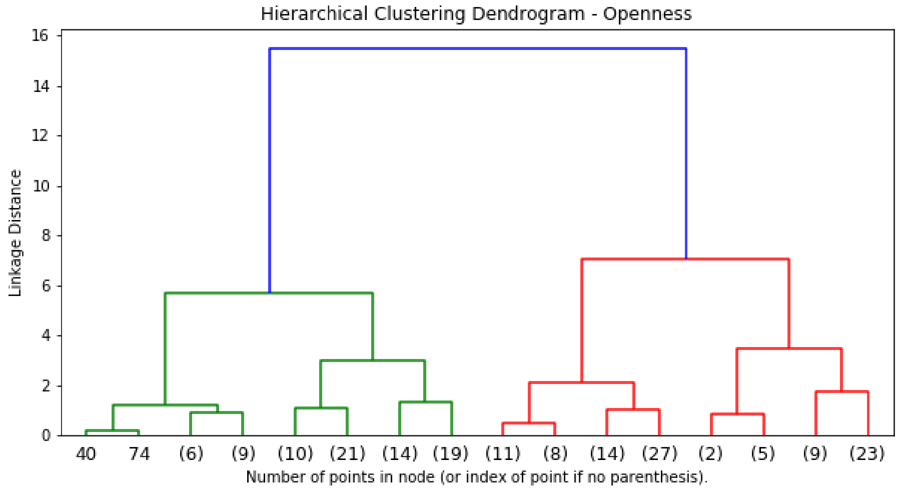


FIGURE B.1: Hierarchical Clustering Dendrogram for trait Openness. Ward linkage function with the top three nodes shown.

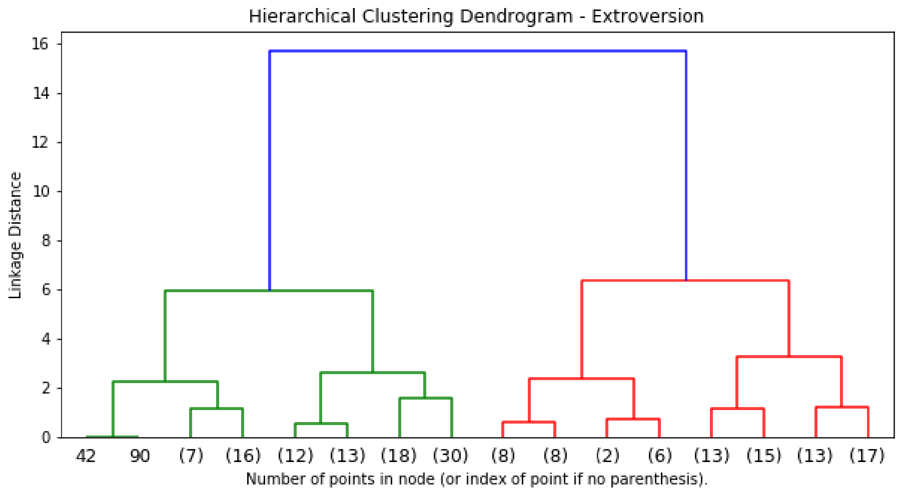


FIGURE B.2: Hierarchical Clustering Dendrogram for trait Extroversion. Ward linkage function with the top three nodes shown.

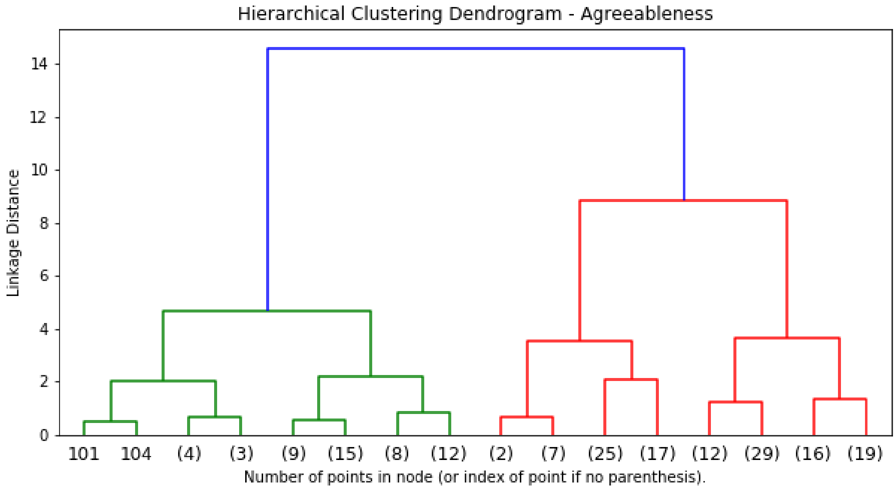


FIGURE B.3: Hierarchical Clustering Dendrogram for trait Agreeableness. Ward linkage function with the top three nodes shown.

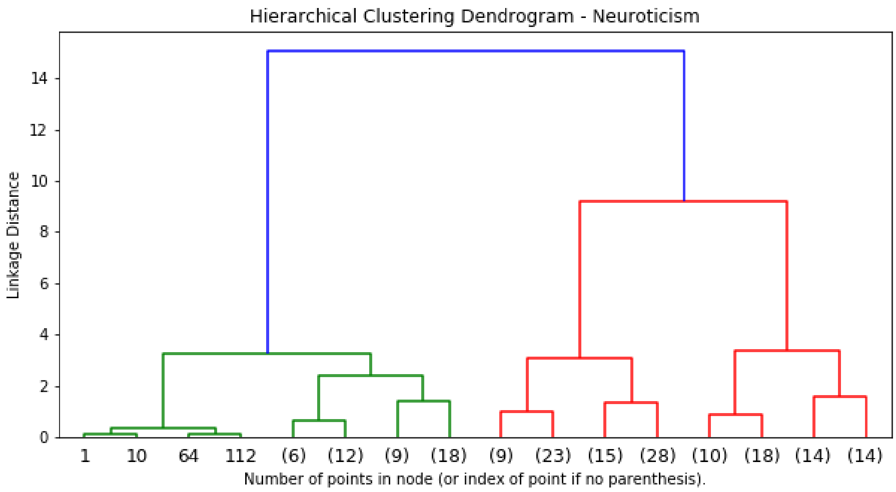


FIGURE B.4: Hierarchical Clustering Dendrogram for trait Neuroticism. Ward linkage function with the top three nodes shown.

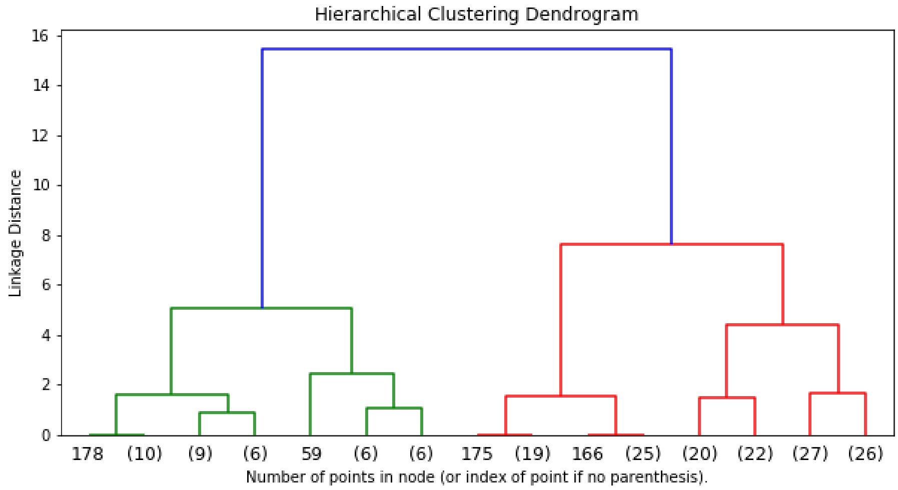


FIGURE B.5: Hierarchical Clustering Dendrogram for trait Narcissism. Ward linkage function with the top three nodes shown.

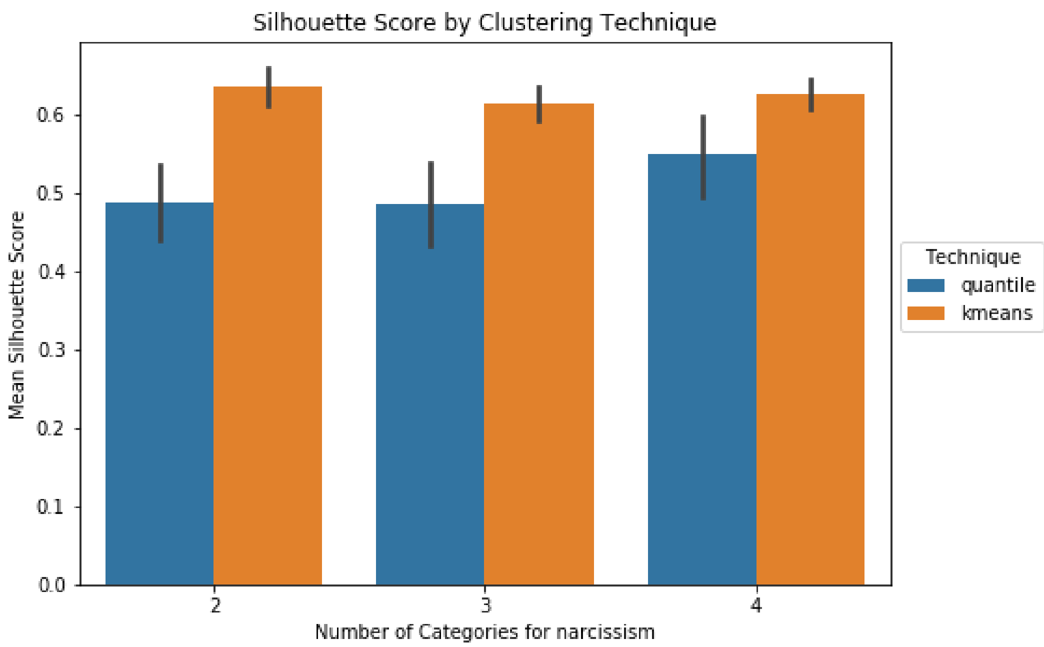


FIGURE B.6: Silhouette score by clustering technique for trait Narcissism.

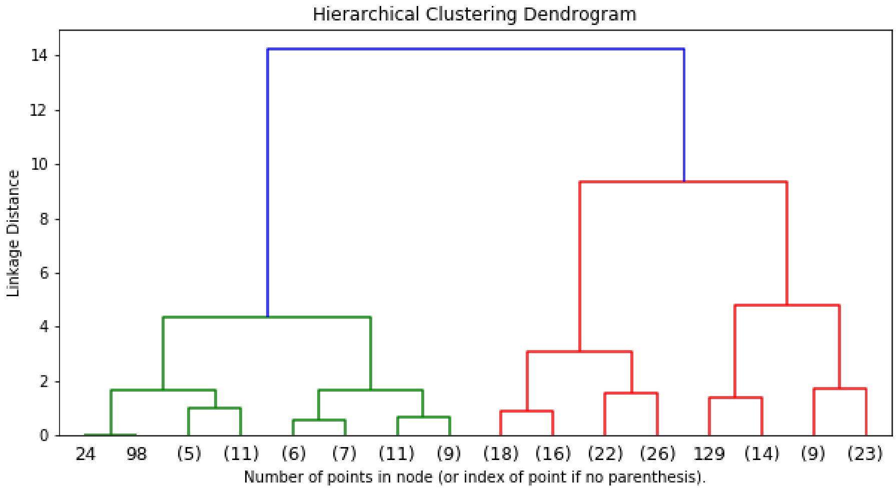


FIGURE B.7: Hierarchical Clustering Dendrogram for trait Self-Esteem. Ward linkage function with the top three nodes shown.

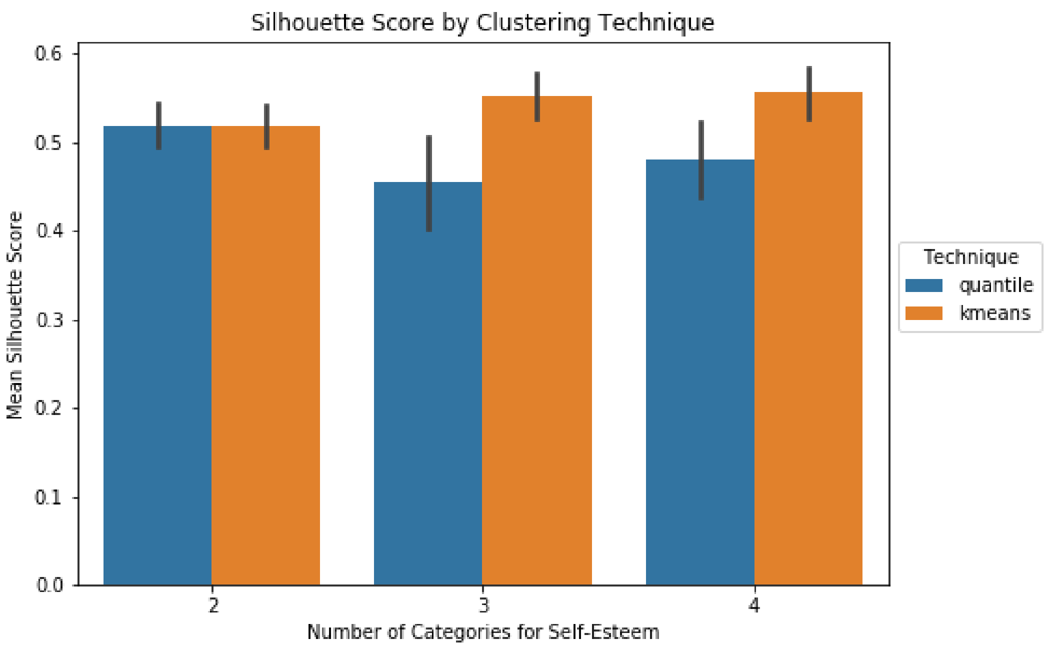


FIGURE B.8: Silhouette score by clustering technique for trait Self-Esteem.

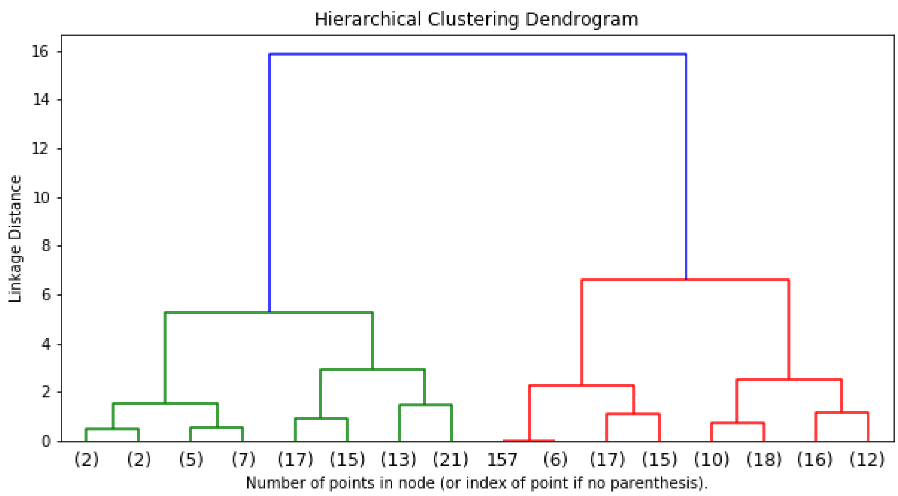


FIGURE B.9: Hierarchical Clustering Dendrogram for Political Inclination. Ward linkage function with the top three nodes shown.

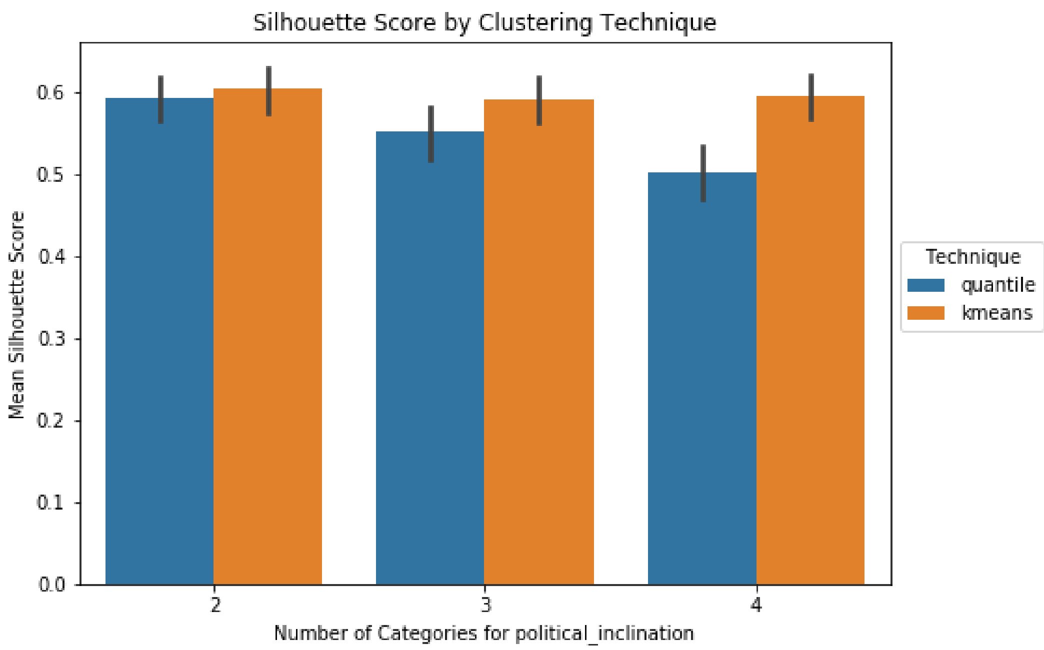


FIGURE B.10: Silhouette score by clustering technique for Political Inclination.