

**MPhil Thesis: The Identification of Gaze Behaviour
associated with decision-making in Surgery: An Eye-
Tracking Study**

Hajra Ashraf
00838165

**Imperial College London
Department of Surgery and Cancer**

MPhil Thesis Submission

The work included in this thesis is entirely original and was conducted by me
as part of an MPhil degree at Imperial College London. All else is
appropriately referenced

The copyright of this thesis rests with the author. Unless otherwise indicated, its contents are licensed under a Creative Commons Attribution-Non Commercial 4.0 International Licence (CC BY-NC). Under this licence, you may copy and redistribute the material in any medium or format. You may also create and distribute modified versions of the work. This is on the condition that: you credit the author and do not use it, or any derivative works, for a commercial purpose. When reusing or sharing this work, ensure you make the licence terms clear to others by naming the licence and linking to the licence text. Where a work has been adapted, you should indicate that the work has been changed and describe those changes. Please seek permission from the copyright holder for uses of this work that are not included in this licence or permitted under UK Copyright Law.

Abstract

Adverse surgical events remain at an unacceptably high level despite multiple global safety initiatives being introduced to practise over the last 15 years. As yet, however there is no conclusive evidence to identify whether physiological markers can be used to predict whether a surgeon will make an error. The aim of this thesis is to identify whether physiological metrics, specifically gaze related metrics can be used to predict whether a surgeon is going to make an error. Thus, my research hypothesis was *gaze metrics can be analysed to provide a greater understanding of perceptual contributions to decision-making in healthcare and aid in the formulation of innovative solutions.*

The research question was initially approached by the undertaking of a literature review, outlining the current knowledge surrounding the uses of eye-tracking technology in healthcare, the findings of which are outlined in chapter one. Next, I undertook two studies recruiting surgeons to perform a series of laparoscopic cholecystectomy tasks on the LapMentor Simbionix trainer. Fatigue severity score (FSS) was calculated prior to the task. Physiological metrics, including gaze tracking, heart rate and skin conductance were measured while they completed the tasks. LightGBM and CatBoost were used to predict the physiological metric most useful in predicting whether a surgeon was about to make an error. My studies indicated that the most predictive feature was the correlation of the skin conductance measure against itself shifted by two time steps (4ms, or 250hz). Skin conductance was twice as likely to successfully predict impending error compared to heart rate and fatigue severity score. Gaze metrics were found to be unimportant thus rendering our hypothesis incorrect. However, when I added gaze features, overall model performance improved by 6.4%. This leads us to surmise that gaze metrics are relevant, however future work would need to establish the importance of this improvement in performance.

This work advances our understanding of physiological changes that occur during decision making and has important implications in the surgical field, as skin conductance could potentially be used as a warning system to indicate when surgeons are at high risk of making an error. The potential for reduction in surgical error rate and improvement in patient safety are important factors to consider.

Table of Contents

No.	Chapter Title	Page
1	Eye-Tracking Technology in Medical Education: A Systematic Review	5
2	Study A - The identification of gaze behaviour associated with making an error during laparoscopic cholecystectomy	32
3	Study B – The identification of gaze behaviour associated with decision making	52
4	Conclusions & Future work	63
5	Appendix	67

CHAPTER 1

Novel Applications of Eye-Tracking Technology in Healthcare Innovation: A Systematic Review

1. Introduction

Eye-Tracking is defined as the activity of studying the way that people's eyes move in order to discover what attracts their attention. Eye-Tracking uses the theory that we move our eyes to bring into high resolution particular portions of our visual field, this allow us to view in fine detail the image occupying our central direction of gaze.[1] Usually we also divert our attention to that particular point in order to focus our attention on the image [2] This mechanism allows us to use eye-tracking technology to follow gaze behavior with a view of giving us information about a person's attention allocation over time. Over the years eye-tracking has proven itself to be a reliable and accurate way of detailing attention allocation in individuals [2].

Studying movements of the eyes dates back to the 1800's and was first described by a university professor by the name of Louis Emile Javal, who noted that reading involves a series of short stops (fixations) and rapid movements of the eye (saccades).[3] Skipping ahead a century, in 1900 Raymond Dodge conducted an experiment where he asked subjects to identify colours whilst executing specific eye movements to identify the type of information picked up by the eye when it moved. Dodge reported "when the eye movement was unbroken, the observer was unable to tell what had been exposed or even that anything at all had broken the black of the perimeter" thus proving that saccadic eye movements do not harvest any cognitively useful information. [4] This landmark study allowed scientists to concentrate on the use of fixations as the most relevant parameter in eye-tracking studies. In the 1950s, Alfred L. Yarbus completed groundbreaking eye tracking research and his 1967 book, *Eye Movements and Vision*, is still used for reference in modern study. Yarbus. et al proved that task given to a subject has a very large influence on the subject's eye movement. [3] Around 1963 David Robinson developed the scleral search coil technique, which allowed researchers and novices to measure eye movements exactly without the need for extensive technical knowledge [5] Initially this method was not compatible with human subjects, and was later adapted for use by human observers by Collewyn, van der Mark, & Jansen in 1975. [5] This method is still popular today and is regularly used in eye-movement research. At around the same time, Cornsweet & Crane developed the dual Purkinje image eye tracker by SRI International, which allowed non-invasive, high precision and low-noise measurements in humans.[5] These technologies have been considerably improved over the last 30 years and today eye tracking technology is more accessible than ever before and is a popular choice for researchers from multiple disciplines.

Eye-tracking methodology has already been used with success in various different fields, most notably within the aviation industry, marketing & research and sports. Within the aviation industry the analysis of eye movements has been used to provide insight into pilot error, pilot fatigue, pilot performance and cognitive load. Studies of the way expert and novices view air traffic has led to the formation of a systematic out the window (OTW) visual scanning pattern for pilots to see and avoid other aircraft [6] The USP of using eye-tracking methodology in advertising is the fact that it eliminates recall error and gives accurate live information about consumer's cognitive engagement [7]. This has allowed for the emergence of tailor branding in a bid to maximise consumer effect. Eye-tracking methodology has been used in the formation of gaze training programmes in sport. "Quiet eye" training within sports has received much attention as multiple studies have shown that athletes who use this training method produce better results than their traditionally trained colleagues. [8]

Eye-tracking methodology has potential for transfer within medical research, with an expanding body of evidence validating its use within healthcare. The aim of this review is to provide a detailed analysis of relevant healthcare literature to clearly outline the ways in which eye-tracking methodology has contributed to healthcare innovation. Various research methods have been compared within eye-tracking methodology to understand what constitutes successful, meaningful eye-tracking results and how these results should be interpreted to allow maximum impact within the healthcare profession.

2. Methods

The PRISMA Guideline was used to undertake the systematic review for this chapter. A search was conducted of PubMed, OVID: EMBASE + PsycINFO, TRIP Database and Science Direct.

2.1 Selection Criteria

The search terms used in database included "Eye-Tracking" OR "Gaze Behaviour" OR "Fixation" and "Eye-Tracking" AND "Health"

Initial search of databases returned 12,504 results after removal of non-human studies and inclusion of full text studies available only. These articles were then de-duplicated to give a total of 11,374 articles for title and abstract review. After review of title +/- abstracts 6,832 articles were removed, and the remaining articles underwent further review to assess for relevance. A total of 127 relevant articles from reference lists of the main articles were also added to this chain for inclusion into the review. A remaining total of 247 articles were transferred to Mendeley© citation manager for full text review based on the following inclusion criteria:

- i. Eye-Tracking to be used as main technique in method of the study
- ii. Must be an empirical study (not a review)

- iii. Eye-Tracking methodology must have been used within an aspect of healthcare

A further 174 articles were excluded from the study as they did not meet the inclusion criteria leaving a total of 73 articles which have been included in this literature review. The last search was carried out in June 2015. Fig.1 shows details of the screening process in flow diagram form.

Total papers included: **73**

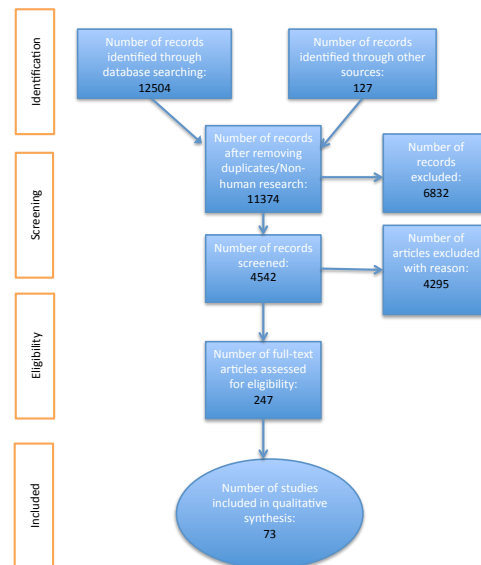


Fig 1: Flow diagram of search strategy

2.2 Data Extraction

Data from the articles was extracted into a standardized pilot data extraction form. Once data was extracted, articles were assessed for common themes which linked them together allowing for the identification of subgroups to allow for systematic review. The following data was extracted from each article:

- **Administrative Details:** Name of author's, Date of publication, Journal
- **Number of participants:** Inclusion and Exclusion criteria of active participants and controls
- **Method:** The use of eye-tracking methodology in the research article
- **Primary outcome measure**
- **Results:** Analysis methods used, results collected and reported outcomes

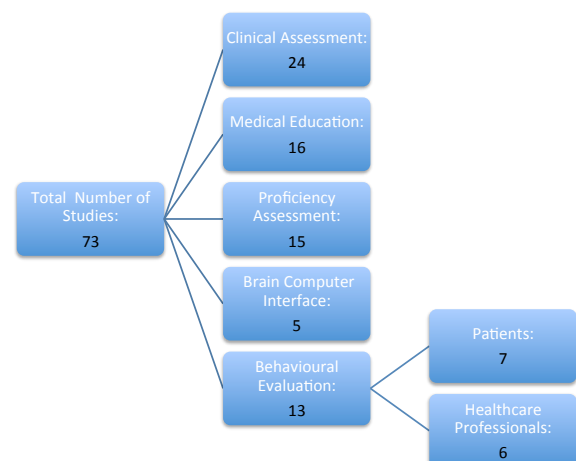


Fig 2: (*adjacent*) outlines how the final count of included articles was split into subgroups.

3. Results

3.1 Clinical Assessment

Ophthalmology

The use of Eye-tracking technology as a means to clinically assess patients has been studied in multiple medical disciplines. Murray et al (2009) used the Tobii x50 (Tobii Technology, Danderyd, Sweden) to assess visual field defects in children and correlated results with the traditional Humphrey Field Analyser (HFA). A total of 29 participants were used in the study - 12 adults with normal fields, 4 children aged < 10 years with normal fields, 8 adults with visual field defect, and 5 children aged < 10 years with suspected visual field defects. Participants were asked to perform 3 eye-tracking tasks as part of the study comprising of a 40-point binocular test and two 41-point tests for each eye. Comparison was then made between the visual field plot results between the eye-tracking tests and HFA tests in adults and between the eye-tracking tests and the clinical assessment in the children with suspected visual field defects. Results showed that for normal adult and child subjects performing the eye-tracking test, the percentage of points in agreement with a healthy visual field was 99.2% and 99.1%, respectively. The percentage of points agreeing with the HFA's screening test in the adult eyes with visual field defects was 89.8% and visual field defects were also correctly identified by the eye-tracking system in the eyes of children with suspected visual field defects [9]. Similarly to Murray et al, Wroblewski et al 2014 identified the use of eye-tracking technology in the assessment of visual fields. Wroblewski et al developed and tested a compact, head-mounted, eye-tracking device called the VirtualEye. This device allowed for assessment of visual fields in two modes: A manual mode where patient response was registered with a mouse click and a visual grasp mode, where the eye tracker sensed change in gaze direction as evidence of target acquisition. Eighty-four participants were recruited and eighty took the VirtualEye test in manual mode and seventy-six in visual grasp mode. Seventy-nine patients successfully completed HFA II test also to allow for comparison. The study showed that VirtualEye head-mounted perimeter had the ability to test the visual field in both manual and visual grasp modes in a manner comparable to standard automated perimetry with Humphrey Field Analyser [10]. Hardiess et al (2010) used eye-tracking methodology to identify the eye and head movements used by homonymous visual field defect (HVFD) patients as a compensatory strategy. Twenty-four patients were recruited into the study, twelve with homonymous visual field defect (HVFD) and twelve normally sighted control participants. All participants completed a simple dot counting (DC) task and a cognitively more demanding comparative visual search (CVS) task while their gaze behaviour was recorded using a mobile eye-tracking system (bright pupil type, model 501 from Applied Science Laboratories, Bedford, USA). Results identified that HVFD patients gaze behaviour showed an increase in gaze movement activity to cope with cognitive demand. [11]

Crabb et al (2014) and Aspinall et al (2015) researched the use of eye-tracking technology as a form of ophthalmological assessment in the elderly population. Aspinall et al (2015) undertook a non-interventional case control study to identify gaze and pupil changes of patients with age-related macular degeneration (AMD) whilst they navigated a simulated walking journey. A total of 57 participants (34 Patients and 23 control) were shown a simulated walking video whilst their gaze behaviour was recorded using a Tobii X 50 eye tracker from Tobii Technology, a dual processor laptop and a 42-inch LCD screen. Visual acuity, fixation count, fixation duration and pupil diameter were recorded and while watching the films, participants were asked to report into a microphone and press a mouse button when the journey was difficult in terms of safety, reduced confidence or uncertainty. Results showed that pupil diameters were significantly greater in the AMD group compared with the control group in both easy and difficult segments of navigation ($p=0.002$) and fixation counts were significantly higher in the AMD group during difficult segments of navigation ($p=0.001$). This allowed Aspinall et al to conclude with a safe degree of surety that fixation count is a marker of difficult navigational environments in patients with AMD. [12]. Crabb et al (2014) attempted to detect age-related neurodegenerative eye disease using eye movement scanpaths. A total of 76 patients (44 Patients and 32 control) were asked to view three unmodified TV and film clips on a computer set which incorporated eye-tracking software (EyeLink 1000, SR Research, Ontario, Canada). Whilst participants viewed the clips the eye-tracking software filtered the data and then generated saccade density maps to identify scanpaths. This study uniquely looked at the scanpaths of patients to identify signatures of vision loss and notably the average sensitivity for correctly identifying a glaucoma patient at a fixed specificity of 90% was 79% (95% CI: 58–86%). [13]. Eye Tracking has also been used to assess reading performance in patients with glaucoma. Smith et al (2014) undertook a study looking at variations between reading duration and eye movements in patients with glaucomatous visual field (VF) loss. This study was a within person study and each person's better eye (least VF damage) was compared to the worse eye (most VF damage). All participants ($n=14$) were fitted for a set of trial frames, which had the appropriate refractive correction for the individual. Participants were then presented with 50 different texts on a screen (one at a time) and were asked to silently read them "as quickly and accurately as possible" whilst one eye was occluded. The eye to be occluded was chosen randomly at the beginning of each task. The screen used had an in build eye-tracking system (Liyama Vision Master PRO 514, Liyama Corporation, Tokyo, Japan) which allowed for gaze behaviour to be recorded whilst each participant was reading the text on screen. Once the participants had read the 50 texts, they were allowed a break before the task was repeated using their alternate eye with 50 different texts. Thus each participant read the same 100 texts but in a randomised order. Results indicated that both reading duration and saccade rate were, on average substantially different in the worse eye when compared to the better eye ($P<0.001$) in 14 patients with glaucoma. Statistical analysis of data showed an "average increase in reading duration in the worse eye compared to the better eye was closely

related to an average decrease in the saccade rate in the worse eye compared to the better eye.” [14].

Paediatrics

Using eye-tracking technology to assess normal and abnormal development of children has also been identified. During infancy the visual system develops rapidly and being able to assess this in preverbal children is vital to identify visual pathology and assess treatment options. Jones et al (2014) explored the use of a remote eye-tracker (Tobii TX120; Tobii Technology AB, Danderyd, Sweden) to identify resolution acuity in infants between the ages of 2 to 12 months. The results of the study were compared with results of the same infants being assessed with the preferential-looking acuity card procedure (current gold standard). A total of 30 infants (aged 2.6 – 12.7 months, with no known visual problems) completed the ACTIVE test, a computer-based system in which a remote eye-tracker tracked infants’ looking responses whilst stimuli are displayed on an LCD screen. Horizontal Gabor patches (stationary sine-wave gratings) were used as stimuli for the test. 1st saccadic movement after stimulus appears and fixation on stimulus were used as an indicator of visual acuity. Jones et al main findings showed that that 90% of acuity estimates using the ACTIVE system were within 90% tolerance limits (based on acuity-card age norms) and measured visual acuity increased with age ($P < 0.001$). [15]. Similarly to Jones et al, Ahtola et al (2014) investigated the visual behaviour of infants, but as opposed to using eye-tracking methodology to identify problems with the visual system, Ahtola used gaze behaviour as a surrogate marker for attention allocation to assess development of infant social skills. Ahtola et al (2014) used eye-tracking methodology to identify new eye tracking metrics for infants’ gaze dynamics. An eye-tracker system was set up (Tobii T120, Tobii Technology AB, Stockholm, Sweden) with a face-distractor competition paradigm. The paradigm used a face or non-face picture as a central stimulus and a geometric shape as a lateral ‘distractor’ stimulus. Data was collected from two samples of infants aged 7-month-old (total $n=45$), and one sample of infants aged 5-month-old ($n=22$). Each infant’s saccadic reaction time from central to peripheral stimulus was measured whilst completing the paradigm. It was found that there is rapid re-engagement of attention on the central stimulus after the initial saccade away from the centre, which would imply a preference for the central stimulus over the lateral stimulus over time. [16] The development of spatial relational memory in children was investigated using eye-tracking methodology by Richmond et al (2015). Richmond et al (2015) showed participants (66 children – 22 nine month olds, 24 eighteen month olds and 20 twenty-seven month olds) three objects and allowed them time to familiarize themselves. Gaze behaviour was tracked using a Tobii XL60 eye tracker (Tobii Technology AB, Stockholm, Sweden) whilst the familiarization process took place. Following familiarization, the participants were once again shown three objects, however two of the objects had switched positions. Results showed that “both 18- and 27-month olds looked significantly longer than would be predicted by chance at the objects that had switched spatial locations; however, 9-month olds did not.” [17]

Psychiatry

Eye-tracking methodology has also been extensively researched with a view for use within the field of psychological assessment. It has been theorized that there are distinct differences in attention allocation between mentally unwell patients and healthy controls. As well as an effect on attention allocation, changes in cognitive function and memory have also been described. [18] Kellough et al (2008) and Isaac et al (2013) both reviewed the changes in attention allocation in patients with a diagnosis of depression. Kellough et al recruited 60 participants (15 major depressive disorder and 45 never been depressed) to their study looking at the time course of selective attention in clinically depressed young adults. It was already known that depressed patients showed a bias for negatively emotional stimuli, Kellough et al (2008) hypothesized that these attentional biases can be sustained over long periods of time. All participants had their line of visual gaze recorded using a remote optics eye tracking system (Model R6, Applied Science Laboratories) while they viewed images from four emotion categories (sad, threat, positive, neutral) for 30 seconds. Results of the study identified that depressed patients spent less time viewing positive images and spent significantly more time viewing dysphoric images, when compared to never depressed patients. It was also noted these biases were maintained over the course of the trial, which would suggest that depressed patients biases for dysphoric information are sustained for substantial periods of time regardless of the presence of other stimuli. [19]. Isaac et al (2013) ran a similar study looking at gaze behaviour of depressed patients whilst viewing emotional stimuli to identify biases for attention allocation. A total of 63 participants - 21 currently depressed (CD), 21 remitted depressed (RD), and 21 healthy controls (HC) were asked to view a slideshow presenting angry, sad, neutral and happy facial expressions whilst their gaze behaviour was recorded using an optics eye tracking system model monocular iView X™ (Sensomotorics Instruments, Version 2.4, Germany). The primary outcome measures analysed were 'first fixation, maintained fixation, relative fixation frequency and glance duration.' Results showed that healthy control patients spent the most time viewing 'happy' images. CD and RD patients showed longer maintained fixations on the three emotional stimuli (happy, sad and angry) compared to the neutral stimulus. Importantly no group differences were found between HC patients and RD patients on glance duration for happy faces, this has important implications with assessment of successful remission of depression.[20]. Beevers et al (2011) employed a similar research method, whereby the relationship between pre-deployment and post-deployment gaze bias for emotional stimuli was investigated in soldiers suffering with Post Traumatic Stress Disorder (PTSD). Beevers et al recruited 139 US Army soldiers who had been deployed to Iraq and had symptoms of PTSD and depression after they had returned home. A remote optics eye-tracking system (Model R6, Applied Science Laboratories - Bedford, Mass.) was used to record line of gaze of participants whilst viewing fearful, sad, happy, and neutral facial expressions once before the soldiers were deployed and again when they had returned home. During their time in Iraq soldiers were asked to periodically report on their levels of war zone stress exposure. Beevers et al found that soldiers with higher war zone exposure showed shorter mean fixation times on fearful facial expressions and this fixation time directly predicted higher PTSD symptom scores. Beevers et al also noted

that longer mean fixation times and longer total fixation times for sad facial expressions predicted higher depressive symptom scores. [21] Social anxiety disorder has been investigated through eye-tracking methodology and gaze avoidance has been recognised as a marker of social anxiety. To better understand this relationship Weeks et al (2013) undertook a study comparing gaze avoidance amongst 20 patients diagnosed with Seasonal Affective Disorder (SAD) and 19 demographically equivalent Non-SAD participants. Participants underwent a social simulation task where they were required to view twenty-six, 12-second videos (13 positive and 13 negative). Participants were unaware of the outcome measures being reviewed in the study and their gaze behaviour was covertly recorded using Gaze Tracker 8.0 (Eye Response Technologies). Results showed that patients diagnosed with SAD showed greater global gaze avoidance in response to both the positive and negative video clips compared to the controls. [22] Chen et al (2013) used behavioural and eye tracking assessments to examine cue detection in prospective memory (PM) of patients with depression (n=19). An SMI iView X-RED Eye Tracking System was used to record eye movements as participants undertook PM tasks (involving a target stimulus that could be present or absent) and an on-going task (target-searching task). PM tasks included cue only and target plus cue, analysis of results indicated that patients with depression had higher number of fixations and longer average fixation durations in the PM task compared to the on-going task. Overall patients with depression performed more poorly than those without depression. [23]. The association between pain [24], mood and age [25] on attention allocation to emotionally positive and negative stimuli have also been investigated. Liossi et al (2014) found that the eye movement behaviour of the patients with chronic pain was characterised by a bias orienting to pain. [24] Isaacowitz et al (2009) used mood induction techniques to identify that 'younger adults demonstrated mood-congruent gaze, looking towards positive faces when in a good mood and at negative faces when in a bad mood. Older adults displayed mood-incongruent positive gaze, looking toward positive and away from negative faces when in a bad mood'. [25]

Smooth pursuit eye-tracking dysfunction (ETD) has been identified in up to 85% of schizophrenic patients and has become one of the most widely investigated biological markers of familial liability to schizophrenia (Shagass et al., 1974; Holzman, 2000). Katsanis et al (2000) undertook twin studies looking at ETD between schizophrenic and healthy twins and found a strong genetic basis for eye-tracking performance. [26]. The measures that account for ETD have been identified as increased root-mean-square-error (RMSE) (Clementz, B.A., Iacono, W.G., Grove, n.d 1996.), increased catch-up and/or intrusive saccades [28], impaired qualitative eye-tracking outcome scores [29] and higher frequency of small anticipatory saccades and leading saccades [30]. Lencer et al (2002) looked at the possibility of analysis of eye tracking dysfunction (ETD) as a marker for the genetic susceptibility to schizophrenia (SZ). A total of 200 participants were recruited to the study – 22 schizophrenia patients, 72 first-degree relatives, 35 other familial relatives and 71 healthy controls. Participants were asked to follow an LED target moving at a constant velocity of 15B/s on a screen whilst their smooth pursuit eye velocity and saccadic frequencies were recorded using an AMTech Eye-

Tracking System (Weinheim, Germany, and SKALAR, Delft, The Netherlands). Eye tracking dysfunction was defined as gain values (eye/target velocity) and saccadic frequencies during smooth pursuit. ETD was shown to be 'significantly worse than it was in non-psychotic families for index patients and relatives with or without a diagnosis of the schizophrenia spectrum in singleton and multiplex families.' [31]. The work of Lencer et al was built upon by Sporn et al (2005), who investigated the differences in eye-tracking dysfunction between family members of childhood onset schizophrenia (COS) patients and adult onset schizophrenia (AOS) patients, using an Ober2 eye-tracking system (Permobil Meditech, Stockholm). In their study, Sporn et al used 154 members of schizophrenic patients' families – these were divided into 70 COS parents, 64 AOS parents and 20 COS siblings and compared their gaze behaviour to that of 161 healthy community control subjects. All participants were asked to take part in a smooth pursuit task, gaze behaviour was recorded whilst participants followed a brightly coloured target presented on a black background, moving horizontally across the screen. Sporn et al concluded that ETD is more prevalent in COS than in AOS, as results showed that RMSE and continuous qualitative score were impaired in COS parents in comparison to community controls. [32]

Cognitive Dysfunction Disorders

Eye-tracking methodology has been employed as a way of investigating the disengagement of attention allocation in patients with cognitive disorders. Hodgson et al (2002) used a complex visual task to identify the variance in gaze shifting strategies between healthy patients and Parkinson's disease (PD) patients. A total of 18 participants were recruited to the study, made up of 8 control patients and 10 idiopathic PD patients. Participants were asked to view a series of pictures showing two arrangements of different coloured balls in two sections of a computer screen. The screen was split into 'workspace' containing the sequence of balls for reference and 'goalspace' containing random sequence of balls. They were asked to plan how they would arrange the balls in the goalspace of the screen to match the balls in the workspace of the screen using the shortest movement sequence possible. The participants also had to verbally indicate the number of moves they needed to make to solve the problem. Scanpaths and number of fixations within each section of the screen were recorded whilst patients were actively problem solving using the EyeLink system (Sensorimotoric systems GmbH). It was found that PD patients made more errors in the task compared to control subjects and did not show any dissociation in the amount of time fixating within the two sections of the screen. Healthy control subjects however, showed a positive correlation between task complexity and time fixating in the 'workspace' section. [33] More recently Crawford et al (2015) undertook a longitudinal study investigating the use of eye-tracking as a biological marker of cognitive impairment in patients with Alzheimer's disease (AD). The study was made up of 25 control subjects and 11 patients that fulfilled the criteria for probable Alzheimer's disease according to the American Psychiatric Association's DSM IV. Saccadic eye movements were recorded using an "ExpressEye" eye tracking system (Optom, Freiburg, Germany) whilst participants completed three different prosaccade tasks. AD patients and control subjects were asked to complete the tasks and were followed up 12 months later, at which

point they were asked to repeat the tasks under the same conditions. It was found that initially AD patients showed slower reaction times compared to control subjects, however interestingly after 12 months both groups showed more accurately directed saccades and an increased speed of reaction times. [34] Impairment of cognitive function in patients with Amyotrophic lateral sclerosis (ALS) was investigated by Keller et al (2015). ALS is severely physically impairing disease and Keller et al used eye-tracking methodology to identify whether the use of eye movements is a viable way of neuropsychologically assessing these patients. The study recruited 48 ALS patients and 32 healthy age and sex matched. The Raven's coloured progressive matrices (CPM) and the D2 test were used in the study, as they have both been well validated methods for measuring executive function. These tests were adapted for use with eye-tracking control and participants were asked to perform these tests using both paper and pencil versions and oculomotor controlled versions. Eye movements were recorded using a portable video-oculography EyeSeeCam device (EyeSeeTec GmbH, Furstenfeldbruck, Germany). The data from the two versions of the tests were compared and it was found that the eye-tracking versions of both the CPM ($p < 0.001$) and the D2-test ($p = 0.024$) were able to distinguish between more and less cognitively impaired patients. [35] The works of Hodgson et al (2002), Crawford et al (2015) and Keller et al (2015) have all shown that eye tracking, as a form of neuropsychological testing is both possible and promising.

A similar problem exists in the assessment of pre-verbal and cognitively impaired children. Neuropsychological testing in this cohort of patients is extremely difficult and thus eye-tracking technology has been investigated as possible solution to this unique problem. As described by Cohen et al in 1988 eye-gaze avoidance is a known phenotype of Fragile X Syndrome (FXS). Hall et al (2009) investigated whether eye-gaze avoidance is associated with high levels of arousal in social situations in children with FXS. Fifty boys and girls with FXS aged between 5 to 20 years were recruited for the study. Subjects were asked to participate in a 25-minute session of intensive social interaction with an unfamiliar experimenter. During this social interaction, cardiovascular activity and eye-gaze avoidance was measured. Hall et al noted that cardiovascular activity was unchanged during the interaction, however eye-gaze avoidance decreased over the course of the 25-minute session. This decrease was attributed to the FXS children becoming more comfortable with the social setting, thus allowing eye-gaze avoidance to be used successfully as a marker of anxiety in FXS children. [36] Jinger et al (2015) used eye-tracking methodology to investigate eye movements of dyslexic children whilst they read Chinese. A total of 62 participants were recruited into the study – made up of 33 fifth graders with dyslexia and 29 age-matched controls. Eye movements were recorded using an EyeLink 2K eye-tracking system whilst the participants read aloud 60 sentences from a computer screen. Time to complete sentence and number of fixations on each word within the sentence were recorded and compared between groups. Dyslexic children exhibited longer and more fixations in across the entire sentence when compared to their controls. It was also noted that as word length increased dyslexic children showed increased number of fixations and gaze durations. This work

identifies eye-tracking methodology as a promising method of evaluating inefficient lexical processing in dyslexic children. [37] Rose et al (2015) investigated the attention and recognition memory for faces in children with Rett syndrome. Rett syndrome is characterized by apraxia and loss of expressive language amongst many other symptoms that impair speech and hand use. Rose et al (2015) recruited 27 patients with Rett Syndrome and 30 (normally developing) controls to their study. A Visual Paired Comparison Paradigm (VPC) was used, where all participants were shown two identical images on a screen, after which the stimuli changed to one new image and the same image shown previously. Preferential viewing on the new image was used to infer recognition. Patients with Rett syndrome did show recognition towards images of both faces and patterns, however their performance was significantly poorer than the normally developing control participants. [38] Autistic Spectrum Disorder (ASD) causes the sufferer to have difficulties attending to and interpreting communicative signals in social situations. Multiple eye-tracking based studies have been conducted investigating aberrant eye gaze mechanisms associated with ASD. Reduced eye gaze duration and lack of specific eye gaze fixation to the eyes and/or mouth have been found to be strong biomarkers of the condition. [39]–[45]

3.2 Medical Education

Clinical practitioners develop their skills over a number of years. Most experiences clinicians have developed their own unique methods of practise over long careers. In order to train competent clinicians it is vital that this knowledge is successfully transferred between experts and novices within their field of practise. Chetwood et al (2012) showed that by reflecting a supervisor's eye-gaze to a trainee while they completed laparoscopic tasks, completion times and number of errors were significantly reduced. [46] The use of tracking eye-movements in order to aid the process of assessment and training in medical education has been proposed multiple times and identified as a promising area of research, with a distinct possibility of future implementation in medical education.

Expert VS Novice

Wood et al (2013) and Bond et al (2014) both looked at eye movement search strategies in the interpretation of electrocardiograms (ECG). Wood et al (2013) aimed to identify the perceptual-cognitive mechanisms used by experts whilst interpreting abnormalities in ECG's. A total number of 20 participants took part in the study split into 10 experts (Emergency medicine consultants) and 10 novices (final year medical students). Participants were asked to make a diagnosis and report their level of diagnostic confidence in 16 ECG traces they were presented with. Participants were fitted with an Applied Science Laboratories (ASL; Bedford, MA) Mobile Eye gaze registration system, which registered their eye movements whilst they scanned the ECG traces. As was hypothesized by Wood et al (2013) experts were twice as accurate and twice as quick at making a diagnosis on the trace and were also much more confident with their decisions. It was found that experts were faster to dwell on the abnormality compared to their novice colleagues. It was also noted that experts concentrated on the surrounding area of the abnormality once it

was located, however scanpath analysis showed novices had a much wider search strategy.[47]. Bond et al (2014) undertook a similar study aiming to identify whether eye-gaze could be used to gain a deeper insight into how experts interpret ECG's. In this study a Tobii X60 eye tracker was used to record the fixations of 21 expert's whilst they interpreted 12 ECG's (3 = easy to interpret, 3 = average and 3 = difficult). These fixations were then used to generate heat maps to identify where the participant looked. As a general principle it was found that leads studied first tended to be studied for the longest. A distinct gaze pattern was also found, with experts most often looking at lead I first followed by leads V2, aVL, V1, II, aVR, V3, rhythm strip, III, aVF, V5, V4 and V6. Lead IV was studied the most on average. [48]

Eye-Tracking as a training method

Occulomotor and eye training as a teaching method was introduced in 2011 by Wilson et al, who undertook a randomized control trial to identify the usability of gaze training for technical laparoscopic skills and performance testing. Wilson et al (2011) recruited 30 trainees with no laparoscopic experience and divided them randomly in one of three groups. The 1st group was gaze trained (GAZE), the 2nd group were movement trained (MOVE) and the 3rd group were allowed to learn on their own (DISCOVERY). Each participant within the three groups was asked to complete ten repetitions of the "eye-hand coordination" task from the LAP Mentor VR laparoscopic surgical simulator. After the series of repetitions to familiarize themselves with the equipment the participants were required to undertake two tasks for data collection a control test and a multitasking transfer test. Eye movements were recorded using an Applied Science Laboratories Mobile Eye gaze registration system (ASL; Bedford, MA), while students completed the tasks. Wilson et al (2011) found that the GAZE group learnt more quickly than the MOVE and DISCOVERY groups. It was concluded that this advantage in performance by the GAZE group was due to differences in gaze control as opposed to other measures e.g. tool control etc. [49]. Vine et al (2012) investigated whether trainees could be taught to adopt the gaze strategies of expert laparoscopic surgeons. A total of 27 novices with no previous laparoscopic training were recruited into the study and randomized into either a discovery-learning (DL) group or a gaze-training (GT) group. All participants then completed 50 learning trials of a laparoscopic training task, while their gaze behaviour was recorded using a Mobile Eye gaze registration system (Applied Science Laboratories, Bedford, MA) to assess gaze control. The GT group completed the trials using a surgery-training template (STT); a novel software designed to highlight key locations on the monitor screen to guide expert-like gaze strategies whereas the DL group had a normal view of the monitor screen. Vine et al (2012) found that the "STT was successful in guiding the GT group to adopt an expert-like gaze strategy (displaying more target-locking fixations). Adopting expert gaze strategies led to an improvement in performance for the GT group, which outperformed the DL group with faster completion times and fewer errors". [50] Similar successes were described by Thiagarajan et al (2014) and Causer et al (2014) in their work using eye-tracking as a form of training. Causer et al used two methods to train surgical residents in 1-handed square knot tying. A total of 20 surgical residents were randomized into the two groups, group 1 was trained using

traditional technical training (TT) and group 2 were trained using an innovative quiet eye training (QET) technique. After the training the TT group were given instructions on how to improve hand movements and the QET group received feedback on their gaze behaviours. All participants then completed simple and complex retention tests under conditions of high and low anxiety whilst wearing SMI eye-tracking glasses (SensoMotoric Instruments (SMI) ETG eye-tracking system -Boston, MA). An improvement in knot tying performance was found in both groups during the post-training test, however the QET group maintained performance when anxiety levels were increased (mean difference: QET, 18%; $P < 0.05$). More efficient gaze and hand movements post training were also noted in the QET group compared to the TT group. [51] The use of eye-tracking as a training method was reviewed by Thiagarajan et al (2014) in view of using it as a rehabilitation technique. Thiagarajan et al (2014) undertook a crossover, interventional experimental design of a single- blinded nature to investigate the effects of oculomotor training (OT) on patients with traumatic brain injury (TBI) as a form of rehabilitation. Versional eye movements of 12 patients with TBI were recorded before and after OMT (fixation, predictable saccades, simulated reading) using the Arrington eye-movement recording system. Thiagarajan et al found that following OMT over 80% of the previously abnormal parameters significantly improved. [52].

Eye-Tracking as a feedback tool

The use of eye-tracking as a feedback and assessment tool and as an adjunct to existing feedback methods has also been shown to be of use within clinical education. Ahmidi et al (2010) showed that adding eye-tracking data to the surgical tool motion data in surgical skill evaluation was of profound benefit. A total of 234 participants were recruited to the study, made up of 95 expert and 139 novices. Participants were asked to find, and touch a given anatomical structure inside the sinus cavity of a cadaver by using an endoscope and a nasal pointer. Motion data and gaze behavior was recorded during the task. It was shown that adding gaze behavior data to the surgical tool motion data improved skill assessment by 13.2% and 5.3% for expert and novice levels, respectively. [53] O'Meara et al (2015) used a similar principle to determine whether the use of eye-tracking technology in addition to video debriefing techniques lead to an improvement in the quality of feedback for students after simulated scenarios. A Quasi-experimental before-and-after study design was adopted and a total of 40 participants made up of nursing and paramedical students from three universities were recruited. Participants completed three 8-minute simulation scenarios of acutely deteriorating patients. Right eye movement was recorded using Tobii™ Eye Tracking Glasses. Participants were later asked to rate whether the use of eye-tracking equipment aided feedback and whether eye-tracking equipment improved learning. On a five point scale 35 (90%) rated the use of eye-tracking equipment to aid feedback highly and 34 (87%) rated eye-tracking equipment improving their learning highly. [54] A similar study was undertaken by Marquard et al (2011) who studied the gaze behaviour of 20 nurses giving medications to 3 patients in a simulated setting. One patient had an id band error and nurses who noticed this error tended to fixate primarily on patients' charts and have a more predictable eye fixation

sequence compared to nurses who failed to notice the error. [55] When combined with video feedback in simulated scenarios gaze behaviour data can provide vital information about attention allocation and decision-making. In the case of medical emergencies, this is decision making process is an important responsibility for all members of the team, however it is a difficult field to study. Szulewski et al (2014) used the combination of eye-tracking technology and first person video review to investigate decision-making and leadership behaviours of team members in a simulated medical emergency. Five physicians made up of a medical student, three emergency medicine residents and one experienced attending emergency medicine physician were recruited into the study. Eye-tracking glasses (Tobii Glasses Eye Tracker, Tobii Technology) were used to obtain data whilst participants took part in a simulated medical emergency scenario. Participants were then interviewed straight after the scenario and shown a first-person recording of their performance with a superimposed gaze indicator, after which independent reviewers studied the videos. The feedback was collated, and it was found that the combination of first-person video and gaze tracking provided information about what the participant looked at, but when combined with interview feedback it was identified that participants may have been looking at a particular area or object without consciously “seeing” it. This phenomenon was closely linked with participants change in focus and indicates that if working memory is preoccupied a participant may see something but in fact be ‘blind’ to what they are looking at. [56] Henneman and colleagues in their study using eye-tracking technology to provide objective evaluative feedback to students after a simulation experience found similar results. Henneman et al (2014) compared the use of verbal debrief, eye tracking debrief and combined verbal and eye tracking debrief in simulation training amongst nursing staff, using a pretest-posttest experimental design. Results showed that safety practises improved most amongst the eye-tracking only group compared to the other two debriefing styles. [57] Khan et al (2012) investigated whether using video feedback of the gaze behaviour of experts would expedite the learning curve of junior surgeons. The study was split into two phases, in phase 1 real-time eye gaze (Tobii X50 eye-tracker) of expert surgeons was recorded whilst performing a laparoscopic cholecystectomy. These videos were then used for phase 2 – where the expert surgeons viewed back their own videos with their gaze behavior being recorded once again. During phase 2 junior surgeons also watched the videos with their gaze behavior being recorded also. When experts watched their own videos, a 55 % overlap of eye gaze was noted, however when junior surgeons watched the videos a 43.8 %overlap ($p < 0.001$) was shown identifying a significant difference in the gaze behavior of the groups. [58]

Changes in gaze behaviour during the learning curve

Junior clinicians must train and actively learn new material to become expert within their specialist fields. Studies have shown that education can, quite quickly affect naïve visual search behaviors as this transition occurs. As such eye-tracking methodology is a useful way of identifying how effective training programmes are. Ghebremichael et al (2014) used eye-tracking methodology to identify how gaze patterns within cognitively salient Areas of Interest (cAOIs) change amongst medical students as they become more familiar with

anatomical landmarks. In this study, Ghebremichael et al were able to identify the gaze behavior changes that occur during active learning and familiarization. Ghebremichael et al (2014) used 31 gross anatomy medical students and tracked their gaze behavior whilst they identified anatomical landmarks on a computer screen. The students were then assigned randomly to a group (A, B, or C) that would repeat this task after one of three course section exams, Back and Limbs, Thorax Abdomen Pelvis, and Head and Neck, respectively once they had completed the Human Gross Anatomy course. Saccadic movements and fixation times on cognitively salient AOIs were measured during the task, both before and after completion of the course. It was found that as students became more familiar with the anatomical structures, fixation times on the white space surrounding AOIs increased, this would suggest that there was an increase in fixation time on cognitively salient AOIs. [59] Soh et al (2014) undertook a similar study looking at the visual search strategies of naïve radiologists transitioning to expert readers after undertaking tutorial type interventions. Soh et al enlisted 14 first-year medical radiation sciences (MRS) students with no previous experience reviewing radiologic images. All participants underwent an image-detection task and were then randomized into two groups – the control group was not given any formal teaching, whilst the experimental group was asked to complete an online e-learning tutorial. Once this was complete, the groups underwent another image detection task whilst their gaze behavior, specifically time to fixation on lesion was tracked using a Tobii X50 (Tobii Technology, Danderyd, Sweden) eye-tracking system. As hypothesized the experimental group showed an improvement in lesion detection overall with a 30% increase in sensitivity ($P = .022$). There was a 49% decrease in mean time to first fixation on the lesion ($P = .016$) in the experimental group and a 45% increase in the mean number of fixations per case ($P = .047$). [60]. Sodergren et al (2011) employed eye-tracking methodology as a way of identifying the effect of training in visual attention behavior. Re-orientation during laparoscopic cholecystectomy was used as the task. Thirty final year medical students with no visual impairment were recruited into a randomized control trial. The students were split into either the control group or intervention group via computer generated randomization techniques. The intervention group was shown a 13-minute video outlining re-orientation strategies in laparoscopic surgery, whilst the control group was given no training. All participants were then shown 12 images of various stages of laparoscopic cholecystectomy and were given the task of interpreting the orientation of the image. Attention as represented by gaze was recorded using a Tobii ET 750 Eye-Tracking system (Tobii Technology, Danderyd, Sweden). Sodergren et al found that the intervention group was more likely to orientate correctly compared to the control group (mean 75.6 versus 56.1 per cent; $P = 0.019$) and that the gallbladder was the most important visual cue used for orientation – as the intervention group focused most on this compared to other structures. [61]

3.3 Proficiency Assessment

Surgery

It is well known from previous research that there is a difference in cognitive and motor skills of surgeons depending on how experienced they are. Control of gaze and attention is vital for the successful execution of these motor skills, and as such eye-tracking methodology can be implemented into use for the assessment of proficiency between clinicians of varying degrees of expertise. Wilson et al (2010) used laparoscopic surgical training tasks to identify gaze control parameters that differentiate novices from experts. A total of 14 surgeons were recruited to the study, 6 novices and 8 experienced. Participants undertook the eye-hand coordination tasks from the LAP Mentor basic skills package whilst their gaze behaviour was recorded using an Applied Science Laboratories Mobile Eye gaze registration system (ASL, Bedford, MA). Comparison of results showed that experienced surgeons spent more time fixating of target location, whilst novice surgeons split their time attention between tracking their tools and target location. [62] Similar methods were used by Harvey et al (2014) and Tien et al (2014) who both used surgical trainees to assess proficiency assessment during surgery. Harvey et al (2014) identified that long quiet eye (QE) duration is directly associated with skill level in sports and that cognitive “slowing down” is a perceptual skill noted amongst expert surgeons. [63] QE duration, hand movement time (MT), fixation location, and fixation duration were recorded in 10 surgeons (3 highly experienced -HE and 7 less experienced - LE) whilst performing a thyroid lobectomy on a cadaver model. Gaze behaviour was tracked using a Mobile Eye tracker (Applied Sciences Laboratory, Bedford, MA). Fixation duration on the ligament of Berry was significantly longer in HE surgeons compared to LE surgeons. QE duration on the recurrent laryngeal nerve was also much longer in HE surgeons compared to LE experienced surgeons (1,647.646245ms VS 872.12691.47ms) [63] Tien et al (2014) used expert and junior surgeons and asked them to complete an open inguinal hernia repair on patients whilst their gaze behaviour was recorded using eye-tracking glasses (SMI Eye Tracking Glasses 2.0, SensoMotoric Instruments, Germany). A total of 9 surgeons were recruited into the study and total of 25 cases were recorded, of which 9 were analysed. Richstone et al (2010) showed that as proficiency in a task improves there is an increase in focused attention and visual representation during performance of said task. This can be measured using eye-tracking methodology using increased fixation rates and a higher proportion of fixation within an area of interest as markers of attention allocation. [64] Tien et al (2014) found that experts had higher dwell times on the sterile field during mesh cutting. It was also noted that experts had higher fixation frequency and dwell time on the operative site during application of mesh compared to closure of the external oblique. [65]

Pathology

Visual diagnostic expertise within pathology has been studied using eye-tracking methodology and the study of gaze behaviour has been shown to be an effective method of identifying the search strategies amongst pathologists with varying degrees of expertise. Tiersma et al (2003) investigated how effectively eye-tracking devices could be used to visualise the scanning patterns of pathologists grading cervical intraepithelial neoplasia (CIN). The eye-tracking system EyeCatcher was used to record the scanning patterns of 5 participants (all pathologists experienced in the grading of CIN) while they

graded two projections of cervical intraepithelial neoplasia. Tiersma et al (2003) were able to distinguish two types of scanning patterns - a scanning type search strategy where the pathologists focused on many points within the image for a short time and a selective type of search where the pathologist limited their search to specific points within the lesion, which they studied for a relatively long time. [66] Krupinski et al (2006) undertook a similar study looking at the human factors associated with the proficiency of pathologists as light microscopists. Krupinski et al (2006) undertook a retrospective cohort study using histopathology glass slide digital images of 20 breast core biopsies. A total of 9 participants were used in the study (split equally into 3 pathologists, 3 pathology residents and 3 medical students) and were asked to look for any abnormalities in the digital images while their gaze behaviour was recorded using an Eye-Tracker SU4000 (Applied Science Laboratories, Bedford, Mass). Using an eye-tracking methodology Krupinski et al were able to demonstrate for the first time that 'when a virtual slide reader initially looks at a virtual slide his or her eyes are very quickly attracted to regions of interest (ROIs) within the slide and that these ROIs are likely to contain diagnostic information' [67] Kundel et al followed up their earlier work in 2008, by undertaking a retrospective cohort study of 400 eye-tracking records collected independently from three institutions. The time required to first fixate a cancer on a mammogram was extracted from the records and it was found that in the first second of viewing 57% of the cancers had a 95% chance of being fixated on. [68]

Radiology

Within the field of radiology there have been multiple studies validating the use of eye-tracking methodology in identifying visual gaze pattern variation between clinicians. Anderson et al (2010) used an inexpensive webcam-based gaze tracking method to identify visual routines (establishing orientation, tracing, indexing, recalling and symmetric comparison) amongst 4 radiology technologists (2 students and 2 well established technologists). Anderson et al found that in images with symmetric anatomical structures where one of the structures is abnormal, the participant typically made a comparison between the two objects to verify the normal and irregular appearance. [69] In the early 1990's Kundel et al identified the use of eye-tracking methodology to detect inter-observer variation in radiology. A simple chest radiograph was used to identify the search strategies of 4 radiologists when presented with a lesion in their central VS peripheral visual fields. Kundel et al (1991) noted that there was an increased visual dwell time in nodule-containing regions (AOI) compared to nodule-free regions and this provided evidence for a covert decision making process to decide whether or not specific features were perceived as a nodule. [70] Since this pioneering study 35 years ago multiple studies have used recognition of abnormalities within the lung fields to help identify and reduce the variability in lung nodule detection amongst radiologists. [71]–[73] Most recently Rubin et al (2015) undertook a study to identify the effectiveness of radiologists' search strategies and nodule recognition using an eye-tracking system (SMI IViewX, RED; Sensomotoric Instruments, Tel-tow, Germany). The remote eye tracker was used to record time-varying gaze paths while 13 radiologists (1–25 years of experience interpreting CT images) interpreted 40 lung CT images

containing a synthetic nodule each. Rubin et al found that detected nodules were within 50 pixels of the closest gaze point for 990 (of a total 992) correct detections. Radiologists searched an average of 26.7% of the lung parenchyma in 3 minutes and 16 seconds and encompassed between 86 - 143 of 157 nodules within their gaze volumes. [74]

Gastroenterology

In gastroenterology – Meining et al (2010) used eye-tracking methodology to compare the perception of images when identifying gastrointestinal lesions by narrow-band imaging (NBI) and standard white light endoscopy (WLE). Eighteen participants with varying degrees of expertise reviewed a total of 23 NBI images of various lesions with the 23 corresponding WLE images. A Tobii eye-tracking system (Tobii X series, Tobii Technology, Danderyd, Sweden) was used to record total time spent on image, time until first fixation of lesion, total number of fixations per image and per lesion and number of fixations until finding the lesion. Results showed that total number of fixations and total time spent on relevant area within the image (AOI) was significantly greater for NBI images compared with WLE images ($P < 0.003$). [75] The visual search strategies of endoscopists was reviewed whilst they performed colonoscopy withdrawal by Almansa et al (2011) The aim of the study was to identify the influence of visual gaze patterns (VGP's) on adenoma detection rate (ADR) – a well-known and validated measure of colonoscopy quality. Visual gaze patterns of 11 endoscopists (all of whom had performed between 64 and 125 colonoscopies) was recorded whilst they watched three videos of colonoscopy withdrawals using an ASL eye tracking system (Applied Science Laboratories Mobile Eye, Bedford, MA). The video screen was divided into 9 equal sections, consisting of one central section and 8 peripheral sections. The percentage of gaze time (GT) in the central vs. peripheral sections was compared using a paired *t test*, after which a *Pearson's test* was used to evaluate VGP with ADR and years of practice. Almansa et al (2011) found that across the study all participants spent more time gazing at the central section of the screen compared to the peripheral (65% vs. 33%, $P < 0.001$) and that 'ADR was significantly associated with increased percentage of central GT ($r = 0.67$, $P = 0.024$) and increased mean GT in the central segment ($r = 0.70$, $P = 0.017$).' [76] Phillips et al (2013) aimed to develop an eye-tracking method to identify abnormalities, which are mobile and changing in size, whereby the method would be applicable to three-dimensional (3D) images. Phillips et al (2013) recruited 6 readers inexperienced in interpretation of endoluminal 3D CT colonography. All readers were shown eight short (<30 seconds) fly-through video clips whilst their gaze behaviour was recorded using a Tobii eye-tracking system (Tobii X50; Tobii Technology, Danderyd, Sweden). Phillips et al found that it was feasible to track gaze behaviour whilst viewing 3D mobile images. Pursuit analysis was used to classify gaze, and these pursuit-based gaze metrics were used to identify varying reader search behaviours for polyp detection. [77] Jaarsma et al (2014) also studied the visual search strategies employed to identify abnormal tissue within the colon. A total of 38 pathologists (13 expert clinical pathologists, 12 pathology residents and 13 medical students) were recruited to the study. Participants were shown 10 microscopic images of colon tissue for 2 seconds each and were asked to diagnose for abnormality whilst

wearing eye-tracking glasses. Results showed that expert pathologists entered relevant areas later in their diagnostic process; intermediate pathologists checked more relevant areas and so took longer to reach their diagnoses whereas novices searched the entire slide in detail. [78]

Other medical fields

Similar methods have been used in multiple other medical fields including orthopaedics, ophthalmology and breast surgery to assess proficiency levels. Wood et al (2013) noted that experts were faster to fixate on the site of a fracture in abnormal x-rays compared to novices. Wood et al (2013) used a total of 30 participants in their study – 10 experts, 10 intermediates and 10 novices at interpreting skeletal x rays. Each participant was shown ten normal and abnormal skeletal radiographs of varying difficulty while their gaze behaviour was recorded using an Applied Science Laboratories (ASL, Bedford, MA) mobile eye gaze registration system. [79]. O'Neill et al (2014) identified the use of eye-tracking techniques to compare the visual gaze behaviours of glaucoma subspecialists and ophthalmology trainees during optic disc examination. A total of 7 glaucoma subspecialists and 23 ophthalmology trainees were asked to review eight glaucomatous optic disc images while their gaze behaviour was recorded using a Tobii eye-tracking system (T120; Tobii Technology, Stockholm, Sweden). Results showed that trainees had no systematic patterns of gaze behaviour regardless of whether morphology was present or not, whereas showed an ordered gaze behaviour strategy and spent longer periods of time fixating on areas most likely to contain pathology. [80] Tourassi et al (2013) investigated 'machine learning for linking image content, human perception, cognition, and error in the diagnostic interpretation of mammograms' [81] Six participants were recruited to the study (3 expert radiologists and 3 radiology residents). All participants wore a head mounted eye-tracker (Applied Science Laboratories - ASL, Bedford, Massachusetts, USA) while they reviewed 20 screening mammograms. Machine learning algorithms were then created to link image content with a) gaze b) gaze and cognition and c) gaze and cognition with diagnostic error. Tourassi et al found that by merging radiologists' gaze metrics and cognitive opinions with computer-extracted image features identified 59% of the readers' diagnostic errors while confirming 97.3% of their correct diagnoses'[81].

3.4a Behavioural Evaluation of Patients

Smoking

Attentional biases can be detected successfully using eye-tracking methodology. These biases have been investigated in relation to alcohol, drugs and cigarettes to better understand how visual cue exposure is related to addiction. Baschnagel et al (2013) used eye-tracking methodology to investigate the drug-use behaviour is maintained among chronic users, in this case smokers. Twenty participants were recruited to the study, which included 12 smokers and 8 non-smokers. The participants were asked to complete a sham experimental task in an office whilst wearing eye-tracking glasses (Positive Science, LLC; Babcock & Pelz, 2004); the participants were unaware of the true objective of the study. During the task the number and

duration of fixations made to smoking cues was recorded. Smokers were found to make distinctly more fixations to the smoking cues compared to the non-smokers, regardless of smoking condition. [82] Avoidance amongst addicts is a known phenomenon and Maynard et al (2014) used eye-tracking methodology to assess avoidance behaviour amongst smokers. Maynard et al (2014) recruited 30 adult dependant smokers and showed them 20 blank, 20 plain and 20 branded packets of cigarettes with familiar and unfamiliar health warnings. Gaze behaviour was recorded using an Eyelink II eye tracker (SR Research Ltd, ON, Canada) to identify the number of fixations within the area of interest (AOI), in this case the health warning on the packet. Maynard et al results showed that regular smokers showed a bias towards fixating on the branding as opposed to the health warning on all three types of packet. [83] Shankleman et al (2015) undertook a study to identify whether branding could affect this avoidance behaviour amongst smokers. A within-subject experiment was conducted involving 30 participants all of whom were not currently smoking and had smoked less than 100 cigarettes within their lifetimes. All participants were shown mock-ups of cigarette packets with their branding in either standardised (plain) or manufacturer-designed (branded) format. Health warnings were present on all packets. Gaze behaviour was recorded using a model TX300 video eye tracker (Tobii Technology AB, Danderyd, Sweden) and results showed that when the packaging was standardised, participants spent longer looking at the health warning label when compared to the branded packaging (mean difference in proportions $\frac{1}{4}$ 0.078, 95% confidence interval 0.049 to 0.106, $p < 0.001$). [84]

Psychological effect of Nicotine

The psychological and cognitive aspects of smoking have also been studied using eye-tracking methodology. In 2008, Gilbert et al investigated the effects of nicotine amongst smokers and non-smokers to determine whether nicotine reduces negative and enhances positive affect as hypothesized by the Situation Trait Adaptive Response (STAR) model. An equal number of smokers and never-smokers were recruited into the study ($n=128$) and were randomized into one of two groups. One group viewed emotionally negative pictures and were asked to focus on the picture stressors. The second group viewed the same stimuli in a two-choice picture attention task that presented the stressor pictures in one visual field and positive pictures in the other visual field. The groups undertook the task twice; all participants were given a nicotine patch during one session and a placebo patch during the other session. The pictures were presented on a monitor interfaced with an Arrington infrared eye-tracking system (Arrington Research Inc., Scottsdale, AZ), which recorded the participants gaze behaviour whilst they viewed the pictures. Results showed that nicotine reduced attentional bias (gaze time) towards negative pictures irrespective of alternative picture valence (positive or neutral). [85] Nathaniel J Wachter undertook a double blind randomized control trial in conjunction with David G. Gilbert in 2012 to identify the effects of nicotine on attention and executive functioning. The study looked at the effect of nicotine on reaction times in an antisaccade task. Participants completed the antisaccade task before and after having a nicotine patch on for 6 hours. The study was double blinded with use of a placebo patch, and sessions took place on counterbalanced days. The antisaccade task involved

36 prosaccade and 36 antisaccade targets randomly intermixed. Gaze behaviour was tracked whilst participants completed the task and reaction times were recorded using an Arrington Research Inc. eye-tracking system (Arrington Research Inc., Scottsdale, AZ). Nicotine was found to have a significant beneficial effect on antisaccade reaction times (RT). [86]

Alcohol

Similar eye-tracking study designs to those used by Baschnagel et al (2013) and Maynard et al (2014) have been used in the assessment of visual bias amongst drug users and the visual bias associated with alcohol consumption. Cairney et al (2004) compared the performance in an antisaccade task of active and ex petrol sniffers. Latency of saccades during the task were used as a tool to identify CNS changes associated with chronic petrol sniffing. Latency of visually guided saccades and antisaccades increased in chronic petrol sniffers, who also had increased antisaccade errors. Cairney et al (2014) suggested this to be indicative of cortical and basal ganglia dysfunction. [87] Rose et al (2013) recruited 64 student social drinkers investigated their attentional bias for alcohol VS non-alcohol drink cues. Participants were asked to choose between simultaneously presented pictures of alcoholic and non-alcoholic drinks whilst their eye movements were recorded using an Applied Science Laboratories (ASL)-6000 remote eye-tracker. Participants were then asked to consume 30 ml of a beverage which was either unchanged or had been altered to taste unpleasant. The task was repeated and visual cues re-assessed. Attentional bias for the alcohol pictures indexed by dwell time ($F = 22.68$, $P < 0.001$), initial fixation ($F = 7.08$, $P = 0.01$) and final fixation ($F = 22.44$, $P < 0.001$) were reduced after participants consumed the altered alcohol, identifying attentional bias as only a partial mediator of alcohol choice. [88]

3.4b Behavioural Evaluation of Healthcare Professionals

Patient safety

Clinical documentation and correct review of medical notes is a vital aspect of patient safety. Currently patients tend to be cared for multidisciplinary teams – which means multiple team members adding to the patients care and writing their entries in the same set of clinical notes. This allows all team members to gain a more complete picture of the patients care, however it can also mean busy physicians have limited time to familiarize themselves with a patient's history. This is why understanding how clinicians obtain information from clinical documents is vital to patient safety. The use of eye-tracking methodology to further study this was undertaken by Brown et al (2014) who recruited 10 participants to read three sets of electronic hospital notes taken from the second day of an adult hospital admission. Gaze behavior was recorded in terms of “fixations” and “glances” to obtain a measure of visual attention, using an ASL Mobile Eye-tracking device (Applied Science Laboratories, Bedford, MA). Results showed that the “Impression and Plan” section of the electronic notes were read slowly, and physicians spent most of their time in this section. “Medication Profile”, “Vital Signs” and “Laboratory Results” were read very quickly and received very little overall attention allocation regardless of whether they contained more content than the

impression and plan. [89] Another aspect of patient safety where eye-tracking methodology has been used is in the assessment of workflow disruptions experienced by surgeons whilst in the operating theatre. Sutton et al (2010) used intra and extracorporeal cameras to record videos of 10 laparoscopic cholecystectomy procedures. These videos were then reviewed and whenever the surgeon's gaze was diverted away from the operation's video display, the event was recorded via time-stamp. Results showed that for every 15 min of operating time, the operating surgeon's attention was disrupted on average 40 times. Although the most common reasons for attention disruptions were necessary (instrument exchange - 38% and downward gaze for extracorporeal work - 28%), the study provided important insight into the areas requiring improvement within the operating room to minimize disruptions and, in theory improve patient outcomes. [90]

Patient - doctor relationship

The patient - doctor relationship is a vital aspect of health care systems across the world. With recent changes empowering patients to take an active role the decision-making process regarding their management being able to understand the dynamics of this relationship is a necessary part of redesigning the role of health care providers. Monatugue et al (2011) used eye-tracking methodology to investigate eye gaze orientation between clinicians and patients during medical encounters. Montague et al (2011) aimed to identify the link between "gaze, clinicians' communication style, and outcomes of the medical encounter"[91] Lag-sequential analysis method was used to evaluate the eye gaze patterns of patients and clinicians in 110-videotapes of medical encounters. Montague et al (2011) results using event-based lag analysis showed that the patients' gaze followed that of clinicians, however this was not true the other way around. [91] The same team undertook a further study in 2012 to identify the differences of eye gaze patterns in three electronic health record interaction styles – a) Technology-centred style b) Human-centred style and c) Mixed interaction style. Videotapes were taken of 100 primary care visits using one of the three interaction styles. These tapes were then reviewed in terms of frequencies and durations of gaze to identify eye gaze behaviours. Similarly to their previous study, Asan et al (2012) used lag-sequential analysis to investigate the association between eye gaze patterns of physicians and patients. Results showed that during technology centred interaction styles, when patients gazed at physicians, the physicians tended to shift their gaze towards the computer and when this shift occurred the patients tended to shift their gaze elsewhere and avoid looking at the physician. It was also noted that interactions using the technology-centred style showed both patients and physician had significantly shorter amount of mutual gaze than other two groups ($p=0.032$; $p=0.015$, respectively). [92]

Re-Orientation during surgery

Operator disorientation during laparoscopic surgery poses a significant risk of injury to patients. Sodergren et al (2010, 2011) have studied the re-orientation techniques used by surgeons using eye-tracking methodology to identify whether effective re-orientation is associated with recognizable patterns in psychophysical behaviour. Sodergren et al (2010) recruited 18

participants made up of 5 surgeons, 2 endoscopists (with varying degrees of endoscopy experience) and a mixture of engineers, medical students and postgraduate researchers. All participants were shown a photograph of everyday objects placed in a box trainer, arranged in a way to resemble the organs found within the peritoneal cavity. Participants were then shown a photograph of the same scene from a different viewpoint as the camera had been moved in one of three possible directions. The participants' task was to re-orientate themselves correctly while gaze behaviour was recorded using a Tobii ET 1750 eye-tracker (Tobii Inc., Sweden). A fixation sequence similarity-based hidden Markov model was used to identify that successful re-orientation strategy relied upon identifying and focusing on a central object within the scene and judging position of its surrounding peripheral objects from there as opposed to using selective, inconsistent feature cues for re-orientation. [93] Comparison of the re-orientation strategies used in natural orifice transluminal endoscopic surgery (NOTES) amongst surgeons of varying levels of experience was also investigated by Sodergren et al (2011) using 21 surgeons made up of 4 attending's, 3 senior residents, and 14 junior residents (all with no surgical NOTES experience). A set of 8 images was shown to each participant, each image was taken from a different perspective during a human NOTES operation helping to induce dis-orientation. A Tobii ET 1750 eye-tracker (Tobii Inc., Sweden) was used to record the gaze behaviour of the participants as they attempted to assess "(1) the configuration of the flexible endoscope within the peritoneal cavity, (2) the image orientation and (3) the orifice used to access the peritoneal cavity". [94] The task was set so that for a participant to have successfully re-orientated they would have had to correctly answer all three questions. Results showed that high-performance participants enlisted a more structured approach to re-orientation as they were found to have lower number of fixations and normalized dwell time per ROI compared to other participants. [94]

3.5 Brain Computer Interface

Gaze sensitive Technology

Brain computer interface is a direct communication between a human brain and an external device, and research within this field dates back to the 1970's. Attempting to use eye movements to control external devices is a novel idea investigated by Noonan et al (2008) who tested the use of a system designed to use eye movements to control an articulated robotic device during minimally invasive surgery. The DaVinci surgical robot was integrated with a binocular eye-tracking system and a 'robotic instrument featuring a long, rigid shaft with an articulated distal tip for minimally invasive interventions'. [95] The system allowed eye fixations of the surgeon in 3D to be converted into commands to direct the robotic probe to where the surgeon desired. [95] Experimental results were successful and showed the ability of this novel eye-tracking system to make possible real-time gaze contingent robot control.

Lahiri et al (2013) investigated how children with Autistic Spectrum Disorder (ASD) who are known to show significant impairments in social communication skills could benefit from the use of a novel "VR-based

interactive system with Gaze-sensitive adaptive response technology”. [96] Eight children, all with ASD diagnoses were recruited into the study and asked to complete tasks on the simulator. The simulator was designed in a way as to facilitate engagement in tasks by using the built in eye-tracking techniques with the aim to improve social communication skills amongst the cohort. Fixation times on the avatar’s face were used to distinguish whether the participant was “not engaged” “engaged” or “highly engaged”. After using the simulator the participants reported that they learned social skills they were previously unaware of, the majority now understood that they should introduce themselves first when speaking to a new friend and that they should look towards the faces of others during a conversation. [96]

Eye-Tracking communication Systems (ETCS)

Eye-Tracking communication Systems (ETCS) are an eye-operated assistive technology used to help people with disabilities to communicate. The use of these systems within children and patients with communication limitations have been studied and results indicate that users have a positive predisposition to the use of ETCSs. Maringelli et al (2013) investigated the use of ETC’s amongst dysarthric ITU patients to identify whether they could improve communication between patients and hospital staff. A total of 38 participants were recruited into the study, consisting of 15 fully conscious medical and surgical patients (all dysarthric due to endotracheal intubation), 8 doctors and 15 nurses. All participants included in the study were provided with an ETC device (MyTobii P10 eye-tracker, Tobii Technology AB, Danderyd, Sweden) and trained on use. A questionnaire was filled out pre and post intervention and answers were compared to identify the whether the ETC devices contributed positively to communication. Results showed that doctors and nurses found that using the ETC device allowed them to better understand their patient’s needs. There was a statistically significant improvement in multiple communication domains amongst patients who also reported reduced feelings of anxiety and dysphoric thought. [97] Similar studies were undertaken by Van Middendorp et al (2014) and Spataro et al (2014). Van Middendorp et al (2014) undertook a longitudinal, prospective, self-controlled cohort study of 6 patients with tetraplegia. Each patient was provided with a Tobii Eyegaze C15 System twice a week for 10 weeks. All participants showed favourable opinions for using the technology. [98] Similarly Spataro et al (2014) provided 30 patients with Amyotrophic lateral sclerosis (ALS) with My Tobii Eye Trackers (iAble-My-Tobii computers, SRLabs, Milano, Italy). A telephone survey outlining usage of the ETC was conducted before and after the patients were provided with the device. Participants found the device most useful for Internet surfing and in communication with relatives/caregiver. [99]

4. Discussion

This chapter review summarises current evidence of eye-tracking methodology contributing to healthcare issues to form innovative solutions. This review has included 73 studies as evidence supporting the use of eye-tracking methodology within various fields of healthcare.

Overall the field of clinical assessment of patients has benefitted the most from studies using eye movement analysis to improve current practise in the fields of ophthalmology, paediatrics, psychiatry and more. Within the field of ophthalmology, all studies reviewed supported the use of eye tracking as a way of detecting visual field defects as accurately as the traditional Humphrey Field Analyser (HFA). [9]–[11]. Important findings with regards to changes in the gaze behaviour of individuals with various ophthalmological diagnoses were also found as a direct result of studies that have adopted eye-tracking methodology. [12], [13] The use of eye-tracking methodology to assess normal and abnormal development of children has been shown to be both possible and of benefit in paediatric care. [15]–[17] Multiple studies included in this review support a growing body of evidence for the use of tracking eye movements to gain insight into psychological disorders, with specific changes in gaze behaviour and attention allocation being found amongst patients with schizophrenia and depression. [19]–[23], [25]–[27], [32] The assessment of cognitively impaired adults and children posed limitations to clinical assessment; eye-tracking methodology has been used in studies to contribute substantially in the formation of innovative assessment methods for this cohort. [33]–[38]

This review also focussed on eye-tracking methodology for use in medical education and proficiency assessment. Training and assessment are vital aspects of producing competent clinicians. Eye-tracking methodology has been used in multiple studies to identify the differences in gaze behaviours of novice and experts. [47], [48] Findings of these studies have the potential to contribute to developing perceptual training programs to train novices effectively in shorter times periods. The use of training as a teaching method has also been identified and studies reviewed showed promising results implicating gaze training methods to improve efficiency of performance and mediate negative effects of anxiety on performance.[49]–[52] Gaze behaviour analysis in conjunction with traditional feedback methods have also proven to be of profound benefit. [53]–[58] The use of eye-tracking as an applicable debriefing technique was shown with uniformly positive results in multiple studies, allowing for a better understanding of decision-making behaviours and cognitive load. Proficiency assessment in surgery, pathology, gastroenterology, orthopaedics, breast surgery, ophthalmology and radiology has been extensively studied using eye-tracking methodology. Studies comparing control of gaze and attention amongst colleagues of varying degrees of experience have provided useful information about successful visual gaze patterns adopted by experts in their prospective fields. [62], [63], [65]–[70], [74], [75], [77]–[81] This novel information impacts heavily on traditional teaching methods and gives rise to an opportunity to adapt teaching methods of the future.

Using eye-tracking methodology to identify attentional bias towards visual cues has been validated as a tool for behavioural evaluation by studies included in this review. [82]–[88]The findings of these studies unanimously suggest eye-tracking metrics, more specifically fixation time, fixation frequency and dwell time can be used as markers of attention allocation. When used to assess attentional bias amongst smokers, alcoholics and drug

addicts, these findings have important implications within health policy and may be used in the formation of novel cognitive behavioural therapies for treatment of addiction.

Eye-tracking methodology has also contributed to aspects of patient safety and patient experience. Studies have shown that using gaze behaviour to better understand how clinicians obtain information from clinical documents, assess workflow disruptions experienced by surgeons whilst in the operating theatre and to assess patient - doctor relationships is both possible and promising. [89]–[92] Within the operating theatre re-orientation strategies in laparoscopic surgery to minimize risk of patients injury was identified. [93], [94] Studies reviewed also showed that the use of Eye-Tracking communication Systems (ETCS) has had positive effects on patient experience, within and outside of hospital. [97]–[99]

Overall this review has outlined relevant healthcare literature in which eye-tracking methodology has contributed to healthcare. Thus far eye-tracking methodology has made important contributions in the improvement of existing clinical techniques. Results of eye movement analysis have also provided us with new information with potential use in many different applications. On the basis of currently available evidence, eye tracking has been shown to be a promising area of research, with applications within multiple aspects of healthcare. Eye-tracking technology is now more readily available than ever before and with progresses in technology allowing for highly portable equipment coupled with its ability to produce high quality quantitative data with minimal setup it is assumed eye-tracking methodology will continue to increase in popularity.

CHAPTER 2

The identification of gaze behaviour associated with making an error during laparoscopic cholecystectomy

Introduction

Adverse surgical events remain at an unacceptably high level despite multiple global safety initiatives being introduced to practise over the last 15 years.[100] The most recent of which, the WHO 'Safe Surgery Saves Lives' campaign accompanied by the global implementation of the WHO surgical safety checklist (2009) has successfully reduced in-hospital 30-day mortality.[101] Errors are inescapable, however having systems in place to prevent them from occurring, remedying them when they do occur, and learning from them has improved overall patient safety in the health care environment.

An error is defined as a failure of planned actions to achieve intended consequences.[102] Errors can be subdivided into two main psychological groups – 'the unwitting deviation of action from intention (slips and lapses) and the departure of planned actions from some satisfactory path towards a desired goal (mistakes).'[102] Slips are related to failure in our memory and lapses are associated with failure in our attention.[102] Error awareness describes the ability to consciously perceive one's own mistakes. Current views suggest that error awareness can be explained by an accumulating evidence account. Whereas aware errors are consciously perceived upon their commission, unaware errors do not cross the threshold of conscious perception and may only be registered retrospectively.[103]

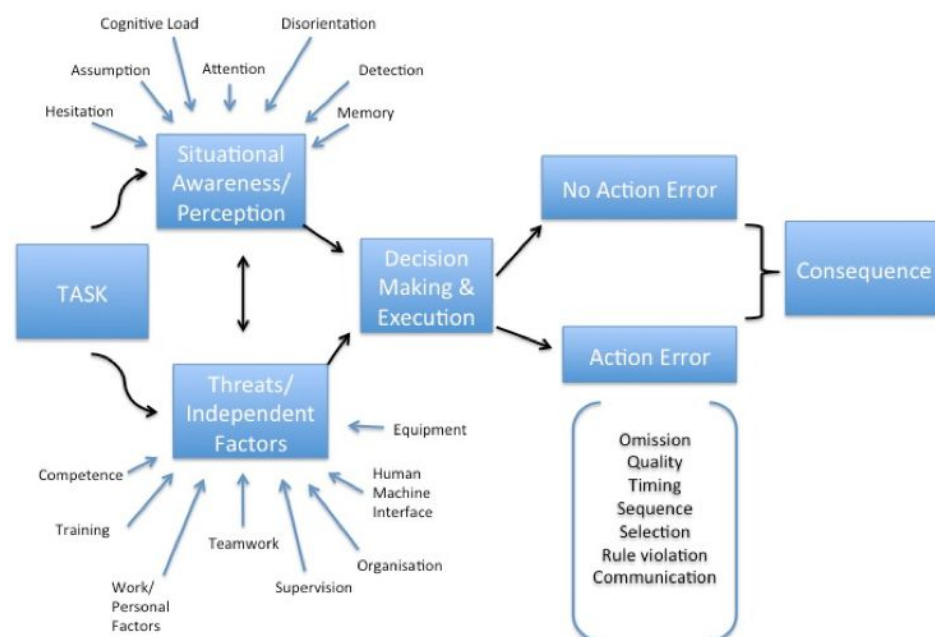


Fig 3. Error Timeline

Neuroimaging has shown that unnoticed (unaware) errors trigger some aspects of error monitoring within the brain, such as anterior cingulate cortex activity and Error-related Negativity (ERN) event related brain potential modulations.[104] Post- error slowing has thus far been used as the primary metric used to examine the behavioural consequences of aware and unaware errors within research. Defined as the behaviourally observed slowing that occurs after the commission of an error (the difference in response time (RT) between post-error trials and post-correct trials.), post-error slowing is thought to reflect an increase in caution of response in trials, that is, a heightening of response thresholds so that response speed is decreased in order to increase the probability of a correct response.[104], [105]

Eye tracking is the process of quantitatively measuring gaze behaviour, using specific metrics, which allow us to identify attentional patterns of individuals during task performance. The use of eye-tracking technology to identify gaze behavior associated with the making of an error is a novel idea. Multiple studies have proven that gaze behaviour changes with various cognitive processes, and the aim of this study is to identify whether there are any detectable changes within gaze patterns that correlate with the making of an aware vs. unaware error. This work consists of two studies to answer this question. In this study I aimed to review physiological metrics which can be used to predict whether a surgeon will make an error. This was done by evaluating metrics recorded while surgeons undertook a simulated laparoscopic cholecystectomy and identifying errors made. These segments along with the various metrics were then used in machine learning algorithms to identify whether there was any correlation between single metrics and the likelihood of a surgeon making an error.

Methods

This study was reviewed and approved by local research ethics committee, reference 15/SW/0356. It was conducted at a single academic surgical centre in London, UK.

1. Participant Selection

All surgeons with a theoretical and practical knowledge of how to perform a laparoscopic cholecystectomy were included in the study. All participants had to have performed a minimum of 10 laparoscopic cholecystectomies on the LAP Mentor Simbionix™ simulator and must have assisted or performed a minimum of 10 in theatre. Informed consent was obtained from all participants and demographics were recorded prior to the experiment. All surgeons were calibrated to a desktop based eye tracking system, data was recorded from initial calibration to end of task.

2. Apparatus

SensoMotoric Instruments (SMI, Berlin, Germany) REDn Scientific™ screen based eye tracking system was used in the experiment. The device uses

infrared light, which is beamed into the eyes whilst a camera records the position of corneal surface reflection relative to the pupil[94]. The software, Experiment Suite Scientific™ then calculates how the infrared eye cameras relate to the image on screen and shows where the eye is pointed. The device is fully automated and allows for head movement compensation. It was attached to the bottom of the LAP Mentor Simbionix™ monitor. The output was a video of the screen of the simulator with a cursor, representing gaze, overlying the images and gaze metrics recorded in proprietary software (BeGaze, SMI, Berlin, Germany). The REDn Scientific™ has a sampling rate of 60 Hz, a gaze accuracy of 0.4° and a spatial resolutions (RMS) of 0.05°.

During the task, basic physiological data was also collected using a Zephyr™ performance systems bioharness. This bioharness was able to monitor HRT - Heart Rate, ECG – Amplitude and RES - Respiration Rate while the participants undertook the task. The bioharness has a sampling rate of 250 Hz (every 4 ms). A measure of skin conductance, also known as the electrodermal response, is the phenomenon that the skin momentarily becomes a better conductor of electricity when either external or internal stimuli occur that are physiologically arousing.[106] The Biopac Systems UK™ EDA100C Electrodermal Activity Amplifier was used in this study. The EDA100C uses a constant voltage (0.5 V) technique to measure skin conductance as it varies with sweat gland (eccrine). The skin conductance response is also known as the electrodermal response. It is the phenomenon of the skin momentarily becoming a better conductor of electricity when external or internal stimuli occur that are physiologically arousing to a person. Arousal is a broad term referring to overall activation and has been found to be a strong predictor of attention and memory.[107]

The study was undertaken using a SimBionix LAP MENTOR™. Laparoscopic Cholecystectomy module 2 was used for all participants during study A and exercise 1: clipping and cutting on a retracted gall bladder was used for study B. The Lap Mentor provides a comprehensive performance report based on objective metrics post simulation. These performance reports were used later in the study for analysis purposes.

3. Procedure

Each participant was given a pre-task questionnaire to fill out. This questionnaire outlined the participant's age, sex, dominant hand, years of training (YoT), the number of simulated laparoscopic cholecystectomies they had completed (LC Sim), self-reported skill level with simulation (Skill sim), the number of laparoscopic cholecystectomies they had assisted in (Ass LC) and the number of laparoscopic cholecystectomies they had completed independently (Ind LC). Each participant also completed a fatigue severity scale (FSS), this 9 item self-report questionnaire has been validated for use in clinical research.[108] The higher the score out of 63 the greater the fatigue severity.

All participants undertook the study in the same room, blinds were closed, and artificial light was used. Participants were asked to complete all paperwork

first. They were then calibrated onto the eye-tracking system, the bioharness was fitted around the chest of the participant and finally 2 x EL507 disposable electrodermal electrodes were attached onto the participants back to measure electrodermal skin response. Participants were then asked to complete the laparoscopic cholecystectomy procedure on the Lap Mentor. Instructions to the participant explained that they had no time limits to complete the procedure, however they should do so in the safest way possible and conduct the procedure as they would on a real patient. After completion of the task, all participants undertook a NASA-TLX questionnaire to measure perceived cognitive workload.

4. Assessment

The laparoscopic cholecystectomy procedure undertaken in this study was separated into 6 segments for ease of analysis. Segment A started at the beginning of the task, until dissection down to Calot's triangle. Segment B started when the clip instrument was selected until the cystic artery was entirely clipped. Segment C started when the clip instrument was selected until the cystic duct was entirely clipped. Segment D started when the scissor instrument was selected until the cystic artery was cut. Segment E started when the scissor instrument was selected and ended when the cystic duct was cut. Segment F started when the first move was made to dissect the gallbladder from the liver bed until the gallbladder was entirely dissected and the procedure came to an end.

Errors were defined using pre-configured metrics from the Lap Mentor performance reports, these include – perforation made to the liver or stomach, applying cautery to a duct, cutting of an unclipped artery, cutting of an unclipped duct, clipping the common bile duct, tearing of a duct with cautery and clipping of the hepatic artery,

Primary eye-tracking metrics recorded and used for analysis included fixations, saccades and pupil diameter. SMI software recorded data from the experiment and wrote it out to a custom formatted text file. There was a header that contained information such as date and time recorded, software version, sample count, calibration and physical experimental setup data. The data consisted of 46 columns of data, excluding timestamp and trial info. Raw X / Y coordinates, pupil diameters, gaze vectors, positions, and other raw data on each eye were recorded at 30 Hz, with an average of 13000 observations per trial.

In addition to the raw data, the software automatically detected higher level gaze events: fixations, saccades, blinks, as well as user interaction events like clicks. Like raw gaze data, blinks, fixations, and saccades were tracked separately for each eye. Fixations had a start time, end time, duration, X/Y location, dispersion, plane, and associated pupil size. Saccades are the quick movements of eyes between fixation points.[109] They have start and end times, and locations, as well as speed, amplitude and acceleration information.

Areas of interest (AOI) were defined to identify the location of the fixations. AOI data was manually created in the form of a series of timestamps associated with the start of the fixation on the AOI. AOI's identified and used for analysis were as follows:

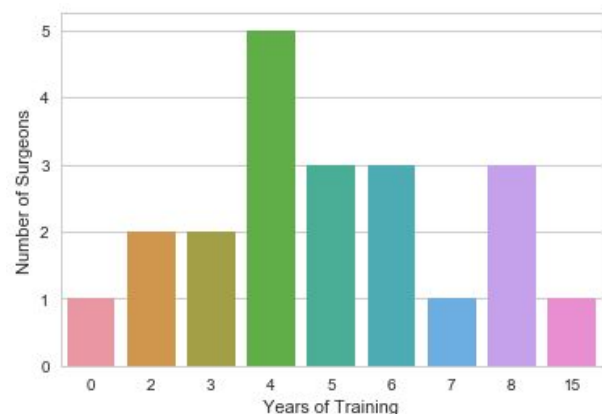
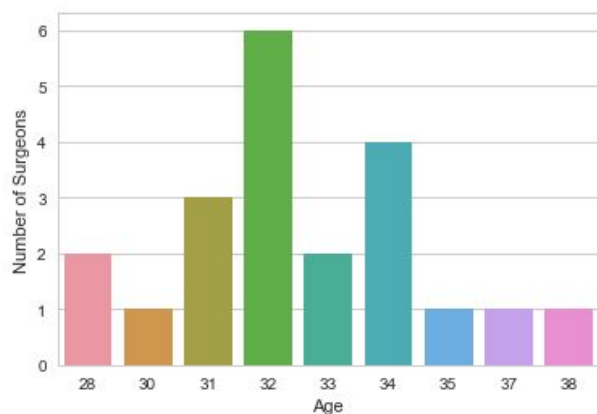
- OFS - Off screen
- BG - Background, when an instrument change was occurring
- L RET - Left retraction instrument
- R DIS - Right dissecting instrument
- R CAUT - Right cauterizing instrument
- R CLIP – Right clipping instrument
- R SCI – Right Scissors
- ST - Stomach
- LIV - Liver
- GB - Gallbladder
- CD - Cystic artery
- CA - Cystic duct

Analysis & Results

A total of 21 surgeons completed the study, of these one participant's data set was incomplete due to equipment failure so the data from 20 participants was used in the final machine learning analysis. Fatigue severity scale results ranged between 20 and 43, with an average score of 32.3. The time needed to complete the surgical task ranged from 7 minutes to 21 minutes, with an average completion time of 13 minutes 48 seconds.

Table 1 outlines basic participant demographics.

Total number of surgeons	21 (5F : 16M)
Average Age	32.5
Self-reported skill rating for simulation	Basic: 5, Average: 14, Good: 2
Dominant hand	19 R : 2 L



I found that years of training had only 35.6% correlation with self-reported skill level. Age correlated even less, at only 16.2%. I considered time to extract gallbladder and found 27.4% correlation with that and skill level. Finally, I considered efficiency of cautery and found 31.3% correlation with that and skill level.

I began by visualizing heart rate data in order to understand which trials (or observations) saw an abnormal heart rate during surgery, where normal heart rate was defined as between 60-100 bpm. A heart rate faster than 100 beats is often associated with stress, anxiety, or fear.[67] It's possible surgeons performing surgeries, even simulated surgeries, may experience higher-than-average heart rate due to the stress of performing a demanding task. Trials 5 and 12 were consistently above 100 beats per minute while trials 1 and 13 saw sudden spikes in heart rate over the course of the procedure.

I continued with correlation observations by generating a table with variables that were correlated above 80%. I then conducted principal component analysis (PCA) in order to convert our set of features which may be correlated into a set of linearly uncorrelated variables (principle components). I split our model into a development set and test set. I chose the minimum number of principal components such that 95% variance (or information) of the model was retained. I found that 11 principal components were chosen. This table demonstrates the linear correlation between each of those 11 principal components and the original 15 features.

Principal Component Analysis on All Trial Data

	ACC - Activity Data	ACC - Lateral Acc Minimum (g) Data	ACC - Lateral Acc Peak (g) Data	ACC - Peak Acceleration (g) Data	ACC - Sagittal Acc Minimum (g) Data	ACC - Sagittal Acc Peak (g) Data	ACC - Vertical Acc Minimum (g) Data	ACC - Vertical Acc Peak (g) Data	BAT - Battery Voltage (V) Data	ECG - Amplitude Data	ECG - Noise Data	POS - Posture Data	RES - Breathing Wave Amplitude (V) Data	RES - Respiration Rate Data	TEM - Skin Temperature Data
0	0.145966	-0.324587	-0.285599	-0.028743	0.468419	0.463881	0.187841	0.113948	0.165942	-0.138381	-0.063405	-0.452112	0.182697	0.139695	0.008383
1	0.371517	-0.302103	-0.145872	-0.001887	-0.223785	-0.238164	0.440321	0.288316	0.411364	-0.048715	-0.003008	0.233659	-0.159770	-0.233230	-0.243913
2	-0.012366	0.169216	0.293149	0.279805	-0.054626	-0.043153	0.135049	0.324754	-0.045161	-0.556117	-0.417810	0.083268	0.393422	0.000520	0.175644
3	0.169986	0.160942	0.309900	0.477243	0.084421	0.114554	-0.063841	-0.147566	0.179492	0.240999	0.226211	-0.143311	0.359745	-0.309226	-0.433887
4	0.257754	0.331619	0.410918	-0.512558	0.066733	0.036554	0.131736	-0.217870	0.268887	-0.013869	-0.250445	-0.060780	-0.054606	0.334740	-0.259747
5	0.165024	0.078355	0.284778	0.186081	0.010701	0.004012	0.228487	0.262381	0.074871	0.123939	0.571913	0.024087	-0.053560	0.477856	0.387138
6	-0.228771	0.141246	0.055547	-0.554154	0.033103	0.025049	0.114923	0.385262	-0.185129	-0.117653	0.434284	-0.042874	0.296266	-0.311386	-0.175765
7	-0.312856	0.039925	0.067003	0.222298	0.069714	0.072122	0.019160	0.477243	-0.194937	0.112219	-0.152331	0.072669	-0.384681	0.288353	-0.546100
8	0.023546	0.132805	0.275779	-0.047779	0.153322	0.140672	0.175989	0.223022	-0.015373	0.410101	-0.289416	-0.184291	-0.303459	-0.492196	0.398718
9	-0.368196	0.081525	0.147179	0.117144	0.099783	0.105727	0.096845	-0.253795	0.336141	-0.529494	0.256606	-0.092660	-0.462787	-0.212702	0.016583
10	-0.653380	-0.185422	0.054706	-0.018322	-0.069758	-0.063092	0.347150	-0.126797	0.348482	0.335803	-0.135065	0.084885	0.327090	0.140784	0.059067



Heart rate data for 21 observations

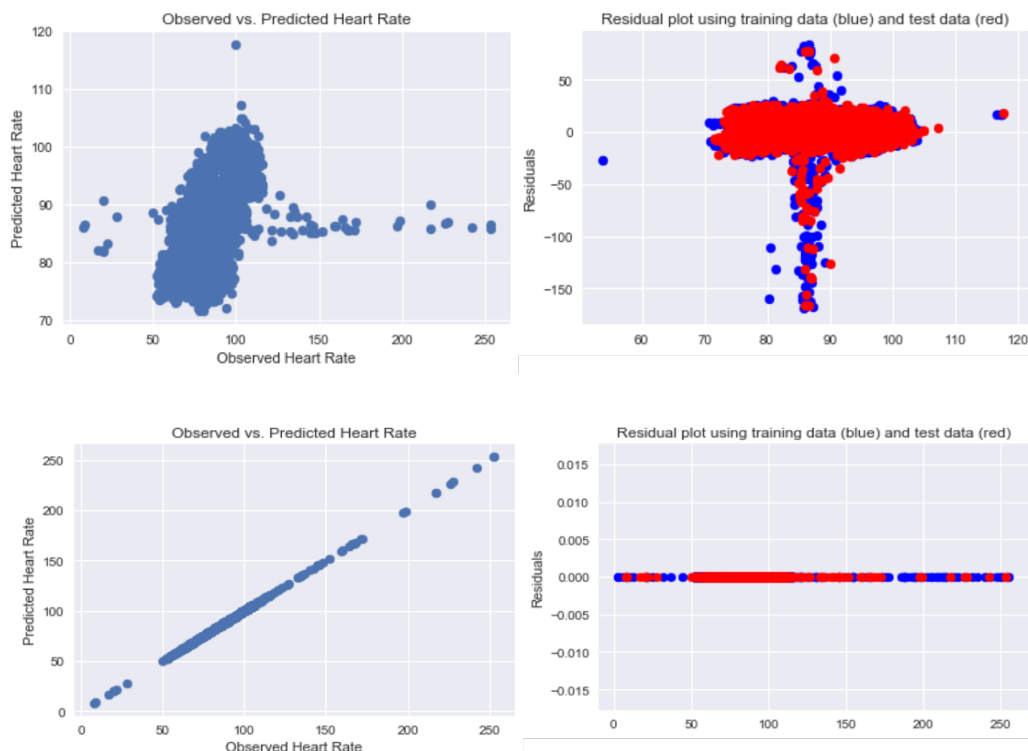
Machine Learning Methods Set Up

1. Predicting Heart Rate

For machine learning methods, I began with the baseline of multiple linear regression and considered all features but 'Timestamp' in order to predict heart rate. I did this to inform the later analysis of instances of surgical error. I randomly separated $\frac{2}{3}$ of the data into a training set and $\frac{1}{3}$ into a test. I emphasized random separation of the training set and test set to prevent patterns from time series observations. The predicted heart rates achieved a mean squared error (MSE) of 169.34 when compared with true heart rate numbers. This is high, so I visualized our observed versus predicted heart rates along with a residual plot using training and test data.



There appeared to be structure in the data, so I applied our PCA.



I achieved MSE=170.86. I trained our multiple linear regression model on a dataset that excluded all of our highly correlated variables. This yielded more interesting results.

My predicted heart rates achieved an MSE of $3e-28$, meaning my model very accurately predicted heart rate. Given the small sample size, it is likely there is an element of overfitting training data.

2. Predicting Simulation Rating

I considered if the features below can be used to predict the skill rating of the surgeons:

- Skill level
- Seconds to extract GB
- Glasses
- Age
- Sex
- YoT
- Efficiency
- Perforations
- Total time

- FSS Score

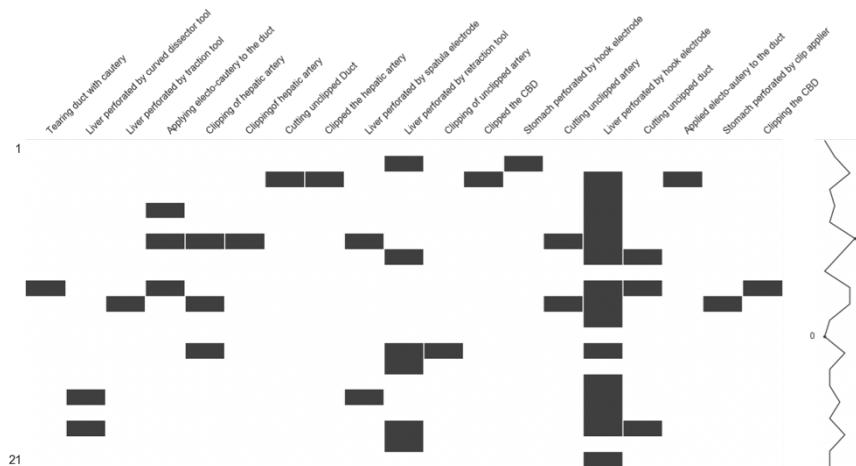
Once again, I separated the data into a training set and test set and conducted principal component analysis to find that 6 components could be used to predict skill rating such that 95% of variance is retained.

	Seconds To Extract GB	Glasses	Age	Sex	YoT	Efficiency	Perforations	Total Time	FSS Score	GFS
0	0.403640	-0.336948	-0.132755	-0.224773	-0.222556	0.406681	-0.277914	0.393883	-0.438385	0.126141
1	0.358107	0.035301	0.587551	-0.019865	0.530764	-0.277618	-0.032972	0.369289	-0.082048	-0.148080
2	0.232781	-0.152253	-0.047459	0.488811	-0.031175	-0.030149	0.423653	0.249560	0.218333	0.624605
3	0.159979	0.596286	-0.147013	-0.423713	-0.199629	-0.252148	0.432317	0.126443	-0.300461	0.145296
4	-0.097016	0.358277	-0.153702	0.340171	0.158044	-0.237366	-0.601135	-0.056071	-0.348893	0.394120
5	0.093168	0.009404	-0.152363	0.601766	-0.231686	-0.138210	0.197157	0.113185	-0.384019	-0.577894

3. Exploratory Analysis



- 2/3 of the surgeons made repeat errors
- The most common mistake was liver perforated by hook electrode (15 surgeons)



The raw measured data (raw gaze, skin conductance and heart rate) needed to be read in from txt and csv files into a Pandas data frame indexed by subject, task, and time. Each file had different time notations (10.11.03 VS 10:11.03 VS 2938810...), which needed to be parsed. Skin conductance data was sometimes denominated in seconds and other times in minutes, requiring conditional processing. Each type of data was also measured at different intervals, HR: 1hz, GZ: 30hz, SK: 250hz. HR data contained a bug where timestamps would overflow as time went from 59 minutes to the next hour (:58, :59, :00, :01...). These were fixed by special casing the parsed data.

For alignment, it was necessary to generate a table of observation meta-statistics of start, end, and durations of each measurement type. Analysis of these statistics showed some issues with raw data alignment, signalling need for synchronization of timestamps. This suggested that gaze data needed to be cropped to given experiment times, while heart rate and skin conductance data needed to be transposed by the experiment start offset. Applying these transformations allowed the three kinds of observation data to be more closely aligned.

As the various types of data had different sampling frequencies, there were many empty values. These missing values were forward filled, as is normal when analysing time series data.

Next, manually annotated data (AOIs, procedure segment, and task error) were associated with the measured data. Error data specifically was used as both labels to be predicted and features. Whether an error occurred in the next 5 seconds were the label, while errors, which occurred in the past, were used as features.

1. Variables to Features

The rows were then grouped into 5 second chunks. Each 5 second chunk of data was associated with a “has_error” boolean variable, indicating whether the chunk contained at least one error.

Each 5 second section contained up to 1,250 rows of data with many columns. TSFRESH was used to derive hundreds of statistical features from the time series data, ranging from min/max values, peak counts and means to more obscure features like time reversal symmetry statistics. The final set of generated features (3,176 different columns) contained many redundant and noisy features, and TSFRESH was used again to filter down to the 48 most predictive features.

In addition to the time series features of observed data, I created a feature to track previous errors that had occurred, decaying by time elapsed. This feature was named “past_error_score”. Each previous error occurrence was considered a score of 1, decaying by 66% every 5 seconds. The previous 10 timesteps were considered, totalling 50 elapsed seconds.

Task procedure data was added as a categorical feature one-hot encoded into 6 columns, one for each segment A ~ F, and zeroed out in the case of being before or between steps. Finally, 13 variables in demographics data were turned into 18 columns of one-hot encoded or scalar data:

- Age
- YoT
- LC Sim
- Ass Sim
- Ind Sim
- FSS
- Perceived cognitive load after short impossible task
- Perceived cognitive load after long task
- Perceived cognitive load after short feasible task
- Feasible or impossible task first feasible
- Feasible or impossible task first impossible
- Sex F
- Sex M
- Dominant hand L
- Dominant hand R

- Self-rated skill sim average
- Self-rated skill sim basic
- Self-rated skill sim good

2. Experiment Set Up: 2 Experiments, 4 conditions

I used two traditional classification models, both in the gradient boosting family: LightGBM and CatBoost. I explored using time series algorithms such as Prophet, but found that most focused on forecasting future numbers, not on classifying time series data.

LightGBM is a gradient boosting machine learning algorithm and framework published by Microsoft™. It is noted for its training speed while maintaining accuracy, using Gradient-based One-side Sampling (GOSS) to reduce the number of data instances to calculate splits for. CatBoost is another gradient boosting algorithm open sourced by Yandex™. It supports more sophisticated categorical variable handling, and an implementation of ordered boosting, meant to prevent certain kinds of overfitting. Finally, CatBoost boasts very fast inference speed, due to the use of symmetric trees for internal decision trees. For real-time error prediction use cases, inferring predictions on real time data streams, fast inference may be of particular interest. Experimentally, CatBoost was about 100x slower than LightGBM to train on a CPU, and 10~20x faster to predict. Both algorithms represent state of the art of gradient boosting algorithms in terms of performance and accuracy and were judged to be a good fit for this project which had a mix of numeric and categorical variables.

The 4 conditions of the experiment were based on different ways to set up and present the dataset to the classifying algorithm.

1. Binary full (0: 3233, 1: 77)
2. Binary subset (0: 77, 1: 77)
3. Multiclass full (0: 3051, 1: 8, 2: 50, 3: 10)
4. Multiclass subset (0: 67, 1: 8, 2: 50, 3: 10)

The binary task used a boolean value of “does an error occur in the next 5 seconds” as the dependent variable, while the multiclass task classified the severity of error (0, 1, 2, 3).

Full datasets represented feeding the algorithms the entire observed data stream. Full datasets included all clips observed, and therefore had many

more rows where no error occurred (50:1 ratio). Subset datasets under sampled the no error class so that observations of “no error” and “has error” were equal, although this still left imbalances in the severity classes for the multiclass task.

I set aside a random 20% of the examples to serve as the test set. I trained the models on the remaining 80% of the data using default hyperparameters and tested on the 20% test set. I originally intended to perform hyperparameter optimization using nested k-fold cross validation, but the number of true positives was so small that I wanted to avoid overfitting and decided to continue with default hyperparameters.

3. Metrics

For each combination of model and dataset, I ran a train/test experiment, taking 7 measures of performance: accuracy, precision, recall, f1, area under receiver operating characteristic curve, Matthews correlation coefficient, and confusion matrix. Recall attempts to test the ability of a model to find all the relevant cases within a dataset.[110] Precision attempts to identify what proportion of positive identifications was actually correct.[110] The formulas used to calculate recall and precision are outlined below in Fig 4. The f1 score is the harmonic average of the precision and recall. An F_1 score reaches its best value at 1 (perfect precision and recall) and worst at 0.[111]

$$recall = \frac{true\ positives}{true\ positives + false\ negatives} \quad precision = \frac{true\ positives}{true\ positives + false\ positives}$$

Fig 4. Recall and Precision formulae ([110])

Overall, LightGBM was fast to train and performed better on unbalanced datasets. Catboost was slightly slower to train, but performed better on experiments with balanced datasets, and achieved the highest performance overall.

Experiment	Model	Dataset	ACC	Precision	Recall	F1	AUC	MCC	CM
lgb_b	LightGBM	binary_dev	96.67	21.43	16.67	18.75	0.7031	0.1722	[[751 11] [15 3]]
cat_b	CatBoost	binary_dev	49.36	3.46	77.78	6.62	0.6323	0.0795	[[371 391] [4 14]]
lgb_bb	LightGBM	binary_dev_bal	64.86	66.67	55.56	60.61	0.693	0.2977	[[14 5] [8 10]]
cat_bb	CatBoost	binary_dev_bal	70.27	73.33	61.11	66.67	0.7003	0.4078	[[15 4] [7 11]]
lgb_m	LightGBM	multiclass_dev	97.44	24.45	24.9	24.67			[[760 1 1 1] [2 0 0 0] [13 0 0 0] [2 0 0 0]]
lgb_mb	LightGBM	multi_dev_bal	47.06	25.43	26.24	25.82			[[10 1 5 1] [1 0 1 0] [6 0 6 1] [1 0 1 0]]
cat_m	CatBoost	multiclass_dev	40.38	25.25	30.35	15.75			[[462 19 228 54] [1 0 0 1] [9 0 4 0] [2 0 0 0]]
cat_mb	CatBoost	multi_dev_bal	41.18	35.92	44.46	33.33			[[8 4 3 2] [0 2 0 0] [4 1 4 4] [1 1 0 0]]

4. Ablations

The previous table describes the performance of different models and different ways of structuring the problem. I wanted to better understand what lead to that performance. For that, I took the best-performing combination of algorithm and dataset (CatBoost + Binary balanced subset) and performed the following ablation tests:

- Numeric observations baseline (numeric): heart rate, skin conductance, gaze raw data
- Full features (+all)

- Full features minus a specific type of feature (-aois, -demo, etc.)
- Baseline plus a specific feature (+aois, +demo, etc.)

Ablation tests involve running the experiment after removing some of the features (or, stated alternatively, by only adding in a few features). If I start with the full feature set and remove a feature, by looking at how much the performance decreases, we can get a sense for how important that feature was. Similarly, by starting with a baseline and adding a feature, by looking at how much the performance increases, we can get a sense for how important that feature is.

The engineered features used were:

- Area of Interest data (aoi's), hand labelled categorical data
- Demographics data (demo), categorical numerical data about subject, including age, experience, sex, task ordering, handed, and self-rated skill.
- Procedure segment (proc_steps), hand labelled data about the current "segment" of the long task procedure
- Past error score (past_err), score representing time elapsed since last error. An error that occurred in the immediately preceding 5 seconds is a score of 1, decaying to 0.

Here I report the metrics. These are not scaled, where the previous table has n100, this table just presents n.

id	acc	auc	f1	mcc	prec	recall	cm
+all	0.8387	0.8375	0.8276	0.6778	0.8571	0.8000	[[14 2] [3 12]]
-proc_steps	0.8387	0.8375	0.8276	0.6778	0.8571	0.8000	[[14 2] [3 12]]
-aois	0.8065	0.8063	0.8000	0.6125	0.8000	0.8000	[[13 3] [3 12]]
-demo	0.7419	0.7417	0.7333	0.4833	0.7333	0.7333	[[12 4] [4 11]]
-past_err	0.7419	0.7417	0.7333	0.4833	0.7333	0.7333	[[12 4] [4 11]]
n+demo	0.7742	0.7729	0.7586	0.5481	0.7857	0.7333	[[13 3] [4 11]]
n+proc_steps	0.7742	0.7708	0.7407	0.5557	0.8333	0.6667	[[14 2] [5 10]]
n+past_err	0.7419	0.7396	0.7143	0.4853	0.7692	0.6667	[[13 3] [5 10]]
n+aois	0.6774	0.6750	0.6429	0.3545	0.6923	0.6000	[[12 4] [6 9]]
numeric	0.7097	0.7042	0.6400	0.4365	0.8000	0.5333	[[14 2] [7 8]]

The +all dataset with every observed and engineered feature performed the best, reaching a maximum F1 of 0.8276 for the balanced dataset.

Removing feature groups one at a time, I found that demographics and previous errors during the experiment were the most predictive features, each accounting for a 0.09 drop in F1 score from the full set. AOI data and procedure step data were less predictive, accounting for a 0.03 and 0 difference in F1, respectively.

Conversely, starting from the baseline, I found that demographics and procedure step data added the most predictive information to the model. This was strange given that procedure data, when removed from the full model, had no effect. I hypothesize that procedure data may hold information that is redundant with the combination of other features, but which by itself can add meaningful information to the model.

It was surprising that AOI data, accounted for such little improvement in the model. Were I to repeat the experiment, the AOI data may be better featured in a finer resolution than 5 second segments.

One challenge was the limited number of errors. Much of the difference in performance metrics was due to changing how I classify a single or very small numbers of errors. For these experiments to be more robust, I would need more examples of errors.

5. Importance

Additionally, I extracted feature importance according to the LightGBM model trained on the binary balanced development set including all engineered features. The table of feature importance's contains all of the features with "importance" of greater than 0. Of 98 features, only 20 had nonzero importance. For this dataset/model combination, the gaze (raw and derived POI) features turned out to be unimportant to classification. The most predictive features are outlined in the table below:

Feature	Importance
conductance__autocorrelation__lag_2	232
heart_rate__fft_coefficient__coeff_60__attr_"abs"	116
fatigue_score	108
heart_rate__ratio_value_number_to_time_series_length	104
past_error_score	90
ldia__number_peaks__n_1	88
conductance__fft_coefficient__coeff_50__attr_"imag"	88
conductance__partial_autocorrelation__lag_1	76
num_real_laparoscopic_cholecystectomies_previously_assisted	50
perceived_cognitive_load_after_short_impossible_task	42
num_real_laparoscopic_cholecystectomies_previously_performed_independently	42
years_of_training	39
perceived_cognitive_load_after_long_task	30
conductance__longest_strike_above_mean	29

self Rated Skill Sim Average	26
num_simulated_laparoscopic_cholecystectomies_previously_completed	19
heart_rate_fft_coefficient_coeff_62_attr_ "abs"	10
perceived_cognitive_load_after_short_feasible_task	8
conductance_autocorrelation_lag_3	4
age	1

Autocorrelation with lag (here calculated with a lag of timestep 2) measures the tendency of this timeseries to correlate with itself. The degree of correlation, or lack of correlation, and sudden changes in correlation over time can be gleaned from this feature. This indicates that the most predictive feature was the correlation of the skin conductance measure against itself shifted by two time steps (the timestep for this data is 4ms, or 250hz).

6. Gaze Metrics

I then focussed more closely on gaze metrics. Five types of gaze events were interspersed together into a single file per subject / task: Fixations, Saccades, Blinks, User Events, Trigger Line Events. Each were associated with a different set of columns, defined in file header. I combined the events for each subject into a single data frame, with a separate column for each event type. Long task offsets were subtracted from each timestamp to align data.

From this I created a new feature gaze_event_score, aligned to long task experiment data and grouped into 5 seconds. The scores were calculated for each type of event (blink, saccade, fixation) at each timestep, and gaze_event_score was the sum of all three.

Formula for gaze event score:

$(N \text{ of detected events in current chunk}) + \text{decay} * (N \text{ of events in previous chunk}) + \text{decay}^2 * (N) \dots$

Reductive tests for each type of gaze event caused a consistent reduction in accuracy and F1 score. Additive tests had a larger effect, with blinks and fixations seeming to cause a slightly larger improvement in accuracy than saccades. The table of feature importance's above contains all of the features with "importance" of greater than 0. For this dataset combination, all of the gaze (raw and derived AOI) features turned out to be unimportant comparative to other features.

CHAPTER 3

The identification of gaze behaviour associated with decision making

Methods

This study set up was identical to the first. This study was undertaken on the same day as the previous study, all participants undertook the 1st study and then moved onto undertaking this study in the same session. In this study I used the same demographic of participants, they undertook a simulated clip and cut task. The instructions were to clip and cut the cystic duct, however the cystic duct was already cut in the simulation, and the only option was incorrect – to cut the cystic artery. This model allowed for us to cause cognitive loading onto the participant as they realized the instructions were impossible to complete. During this time of active decision making physiological metrics were recorded. Participants were randomized into whether they did the impossible task or the feasible first. This was recorded and was later used during analysis.

Analysis

I looked at start and end times, impossible task outcomes, and demographics. I, then cropped out beginning and ending offsets from observed task data so that aligned data began uniformly at 0ms. I removed outliers (< 0.05 quantile, > 0.95 quantile) and centered distribution around 0 to make variance easier to compare.

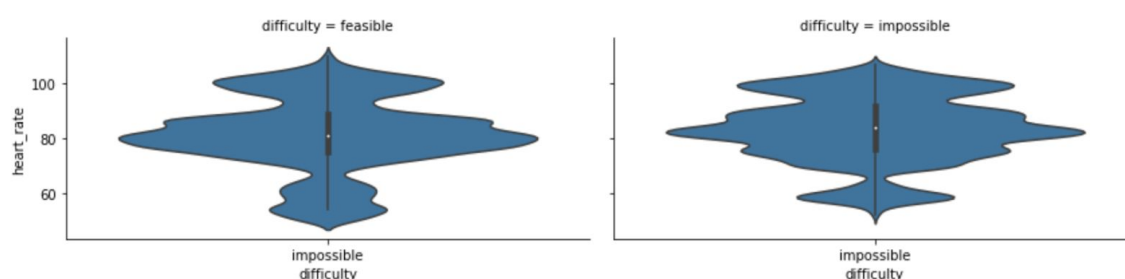
Three conditions used for comparative analysis were as follows:

- a) Impossible tasks vs. feasible tasks (difficulty)
- b) Correctly finished impossible task vs. incorrectly finished (outcome)
- c) Did impossible first vs. did feasible first (order)

1. Heart rate

I completed basic visualization of HR and conductance by subject, to search for outliers.

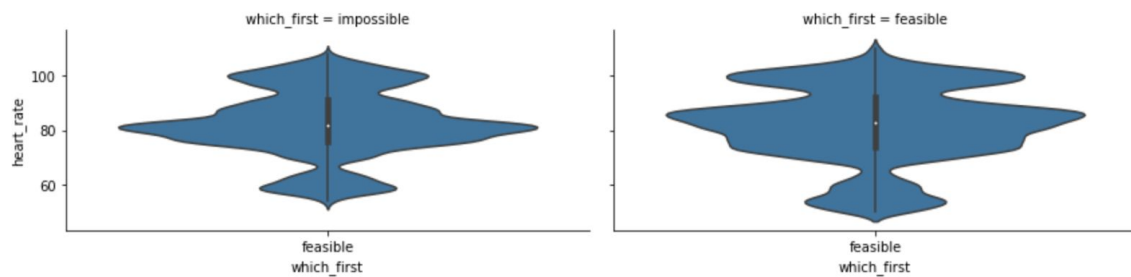
a) HR across difficulty:



	mean	std	min	max
difficulty				
feasible	81.410375	13.652241	50.0	110.0
impossible	83.453118	12.221722	52.0	107.0

Overall the range was very similar, with just slightly higher mean HR for the impossible task, but higher variance for feasible task.

b) HR across outcomes (correct vs. incorrect)

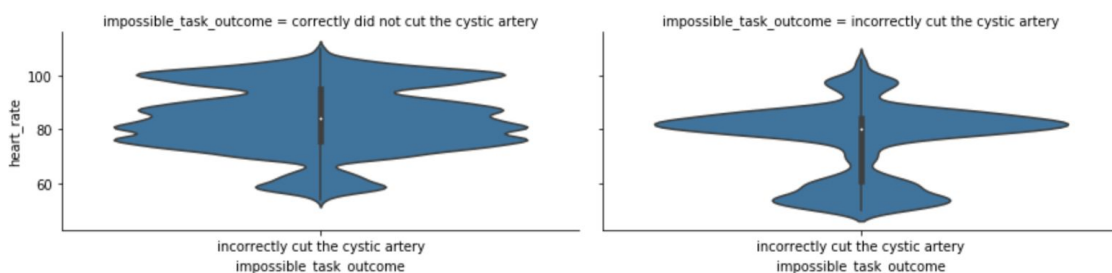


	mean	std	min	max
impossible_task_outcome				
correctly did not cut the cystic artery	84.243334	12.164621	54.0	110.0
incorrectly cut the cystic artery	75.788650	13.831323	50.0	106.0

Those who finished the impossible task correctly had higher mean HR.

c) HR grouped by ordering of tasks

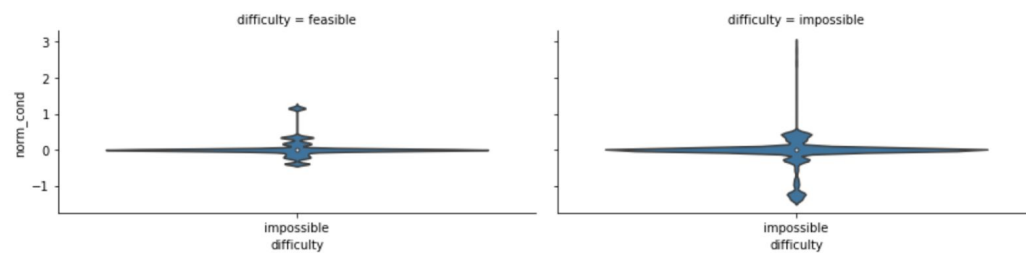
Little difference was found across this split. Slightly higher variance was found when feasible task was completed first.



	mean	std	min	max
which_first				
feasible	82.191920	13.881583	50.0	110.0
impossible	82.567798	12.170415	54.0	108.0

2. Skin conductance

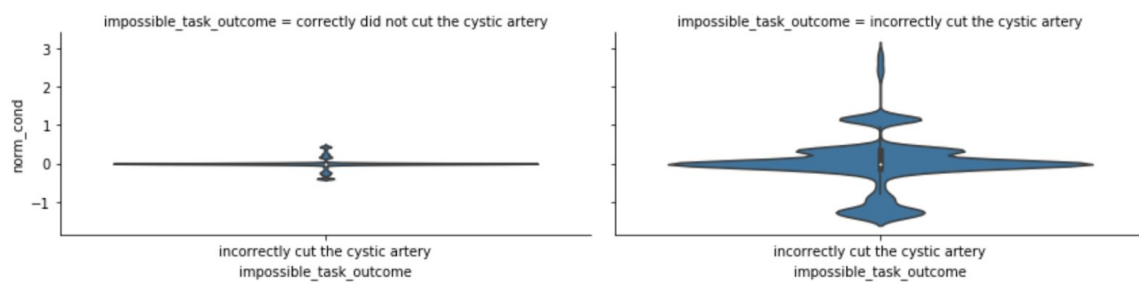
a) Skin conductance by difficulty



	mean	std	min	max
difficulty				
feasible	0.039441	0.254066	-0.400717	1.230209
impossible	-0.051212	0.448006	-1.409561	2.967761

Variance is higher for impossible task

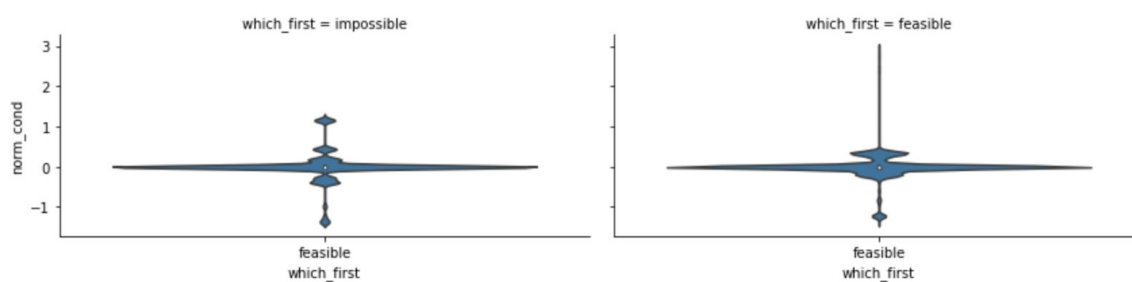
b) Skin conductance by outcome



	mean	std	min	max
impossible_task_outcome				
correctly did not cut the cystic artery	-1.024244e-15	0.134524	-0.400717	0.482794
incorrectly cut the cystic artery	7.620971e-15	0.692387	-1.409561	2.967761

Variance is much bigger for subjects with incorrect outcomes

c) Skin conductance by order

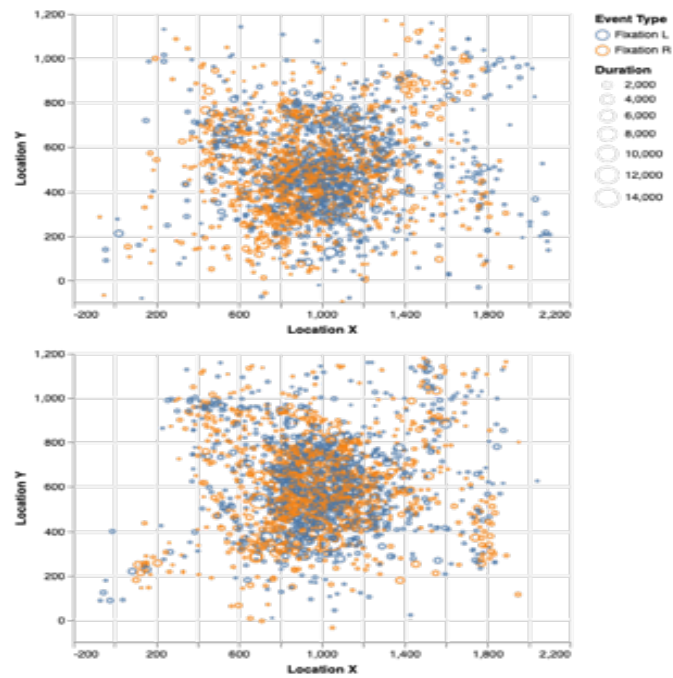


	mean	std	min	max
which_first				
feasible	-7.278979e-16	0.354102	-1.390149	2.967761
impossible	2.310837e-15	0.355262	-1.409561	1.230209

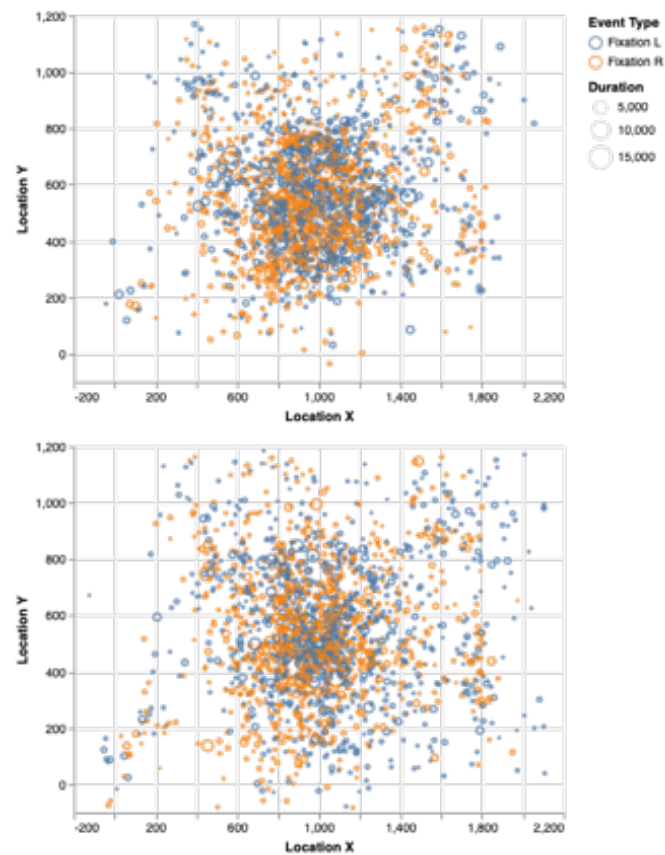
3. Gaze

Visualized gaze distributions (for fixations) across conditions were as follows, but a clear pattern did not emerge.

- Gaze (Fixations) Difficulty - The top chart shows fixations for impossible tasks. The bottom chart shows fixations for feasible tasks.

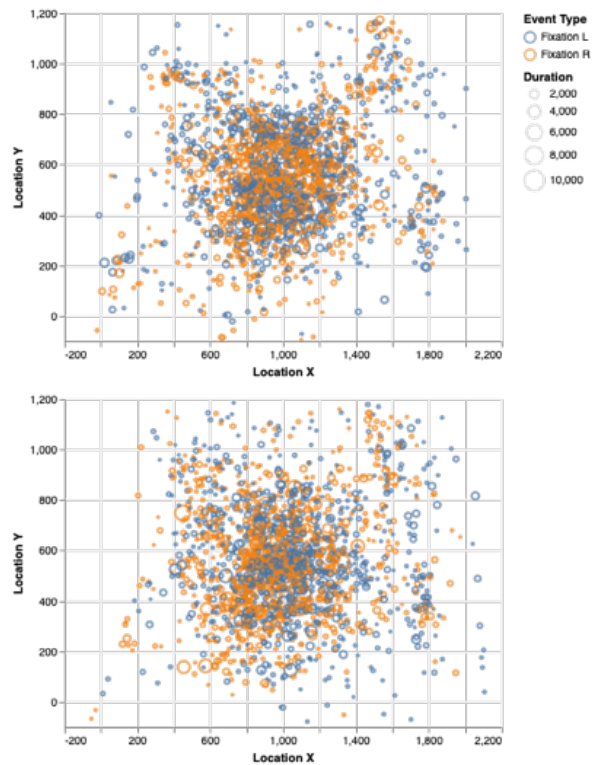


- b) Gaze (Fixations) by outcome - The top chart shows fixations where the person did the correct thing when faced with an impossible task. The bottom chart shows fixations where the person did the incorrect thing and cut the cystic artery.



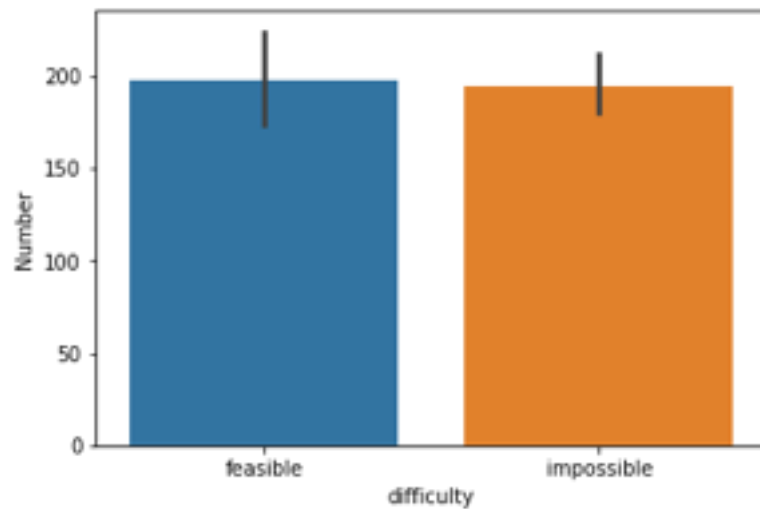
- c) Gaze (Fixations) by order – The top chart shows fixations for people who

did the impossible task first (top) and people who did the feasible task first (bottom).

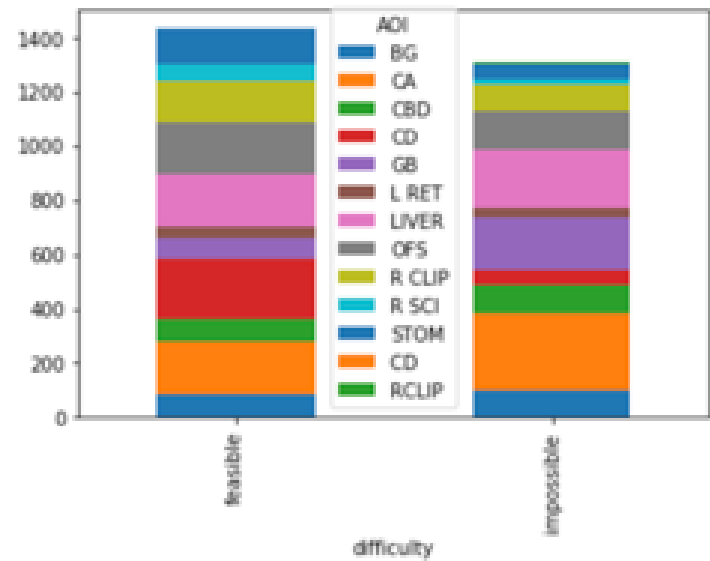


I also reviewed whether the average total number of gaze events were different for any of these conditions.

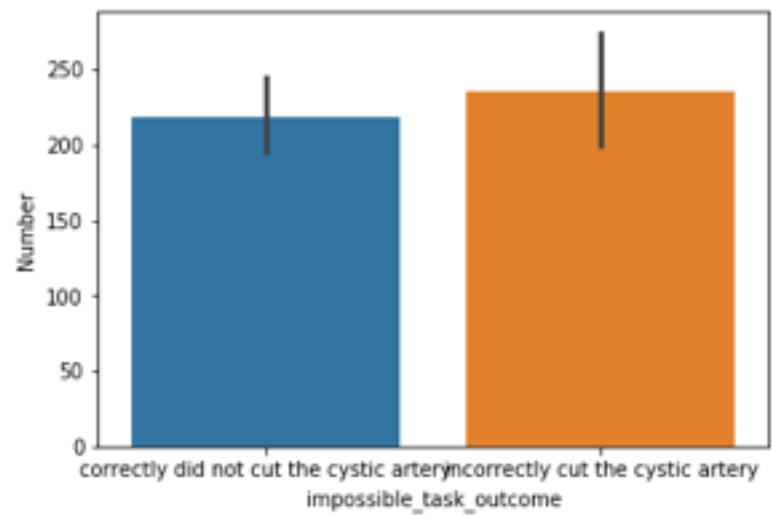
a) Fixations
Difficulty -
Feasible: 197.15
Impossible: 194.65



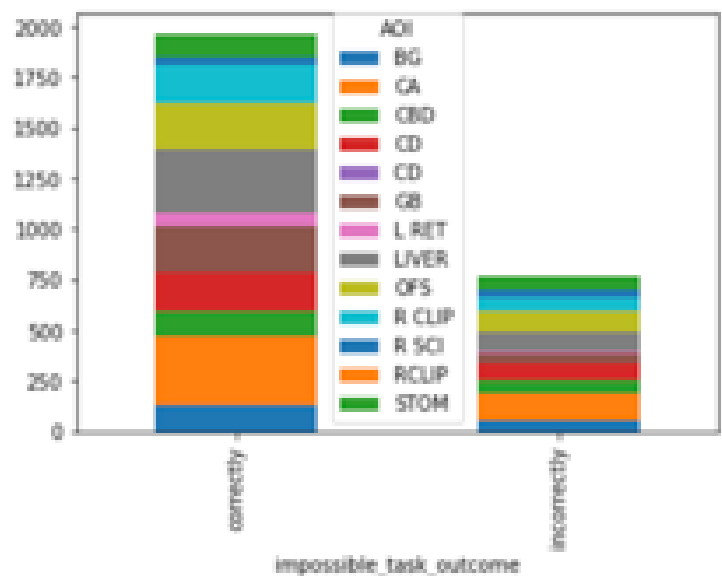
AOI fixation breakdown for difficulty:



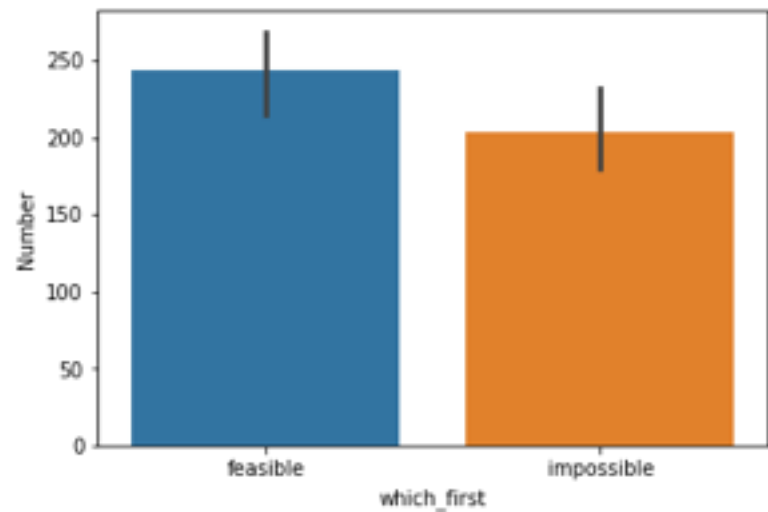
Outcome -
Feasible: 218.36
Impossible: 235.17



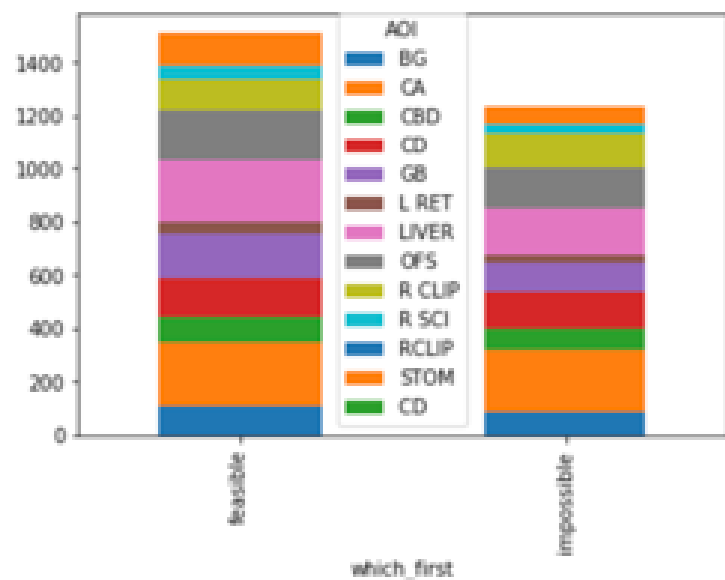
AOI fixation breakdown for outcome:



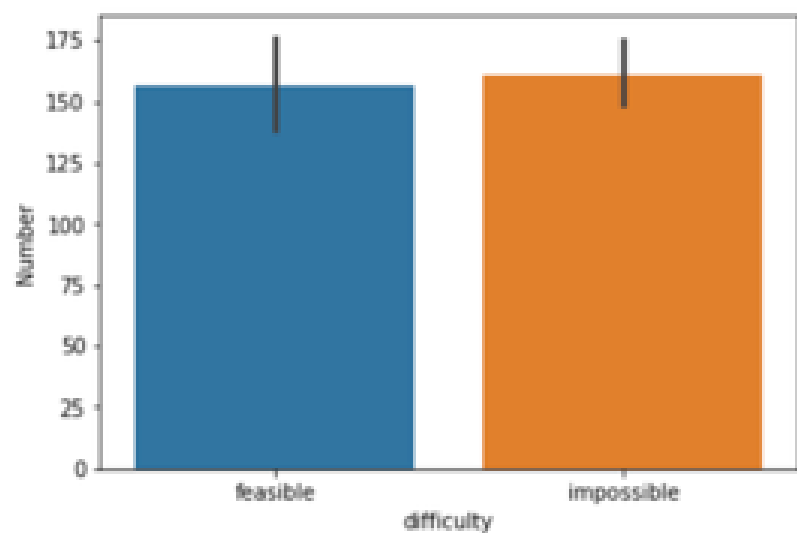
Order-
Feasible: 243.9
Impossible: 202.9



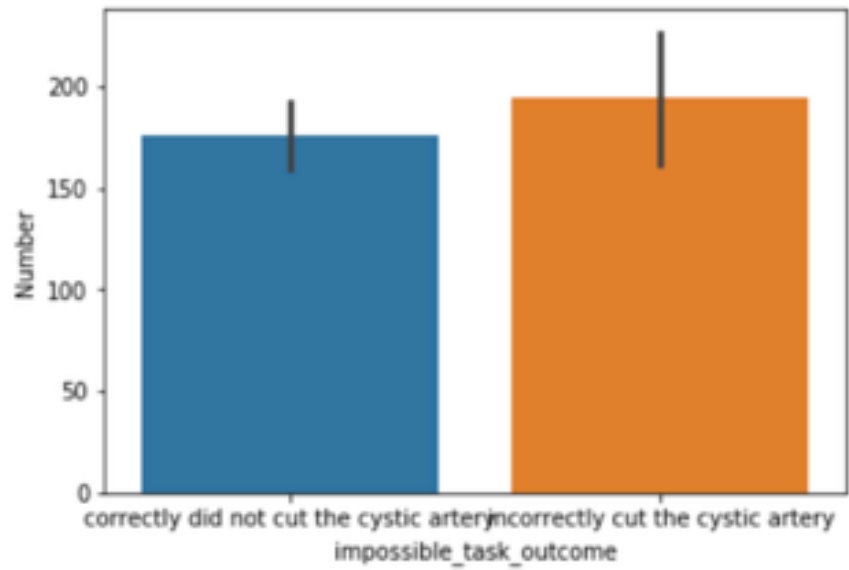
AOI fixation breakdown for order:



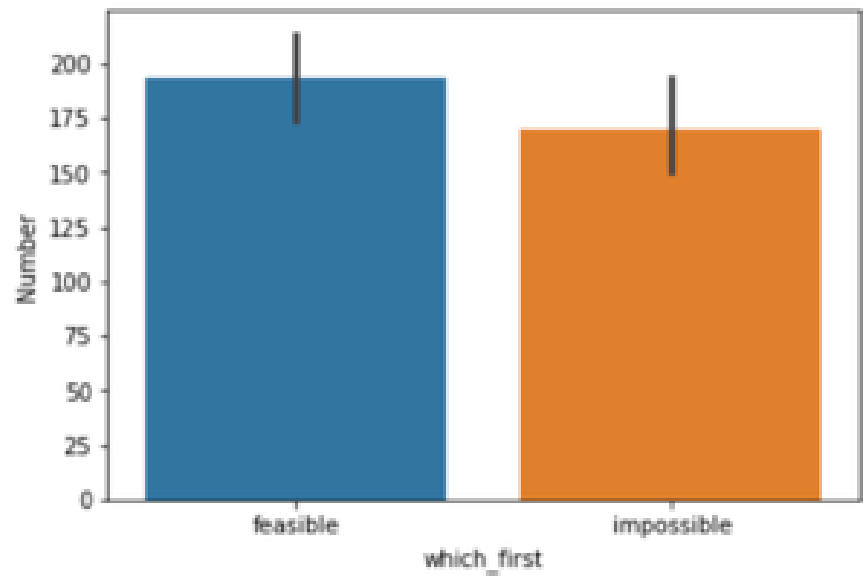
b) Saccades
Difficulty -
Feasible: 156.65
Impossible: 160.95



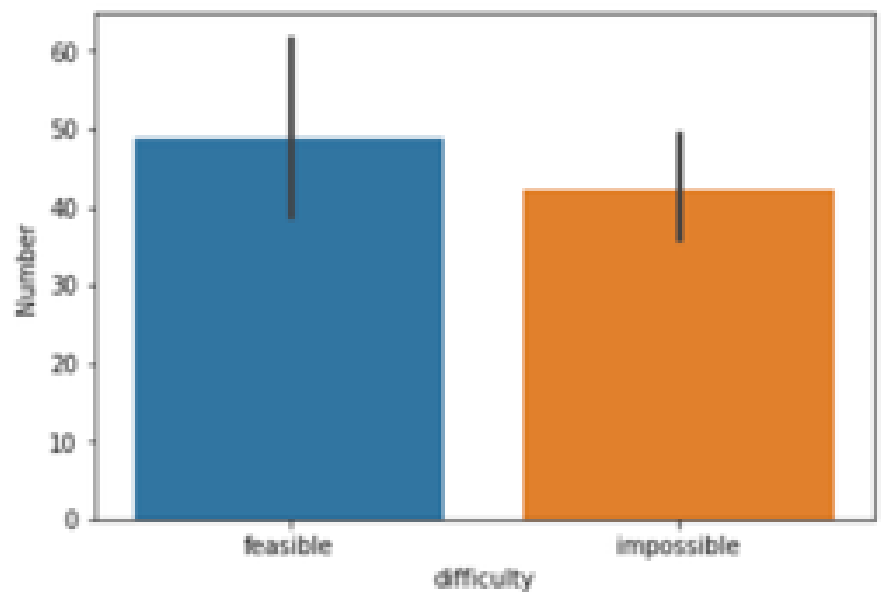
Outcome -
Feasible: 175.85
Impossible: 194.5



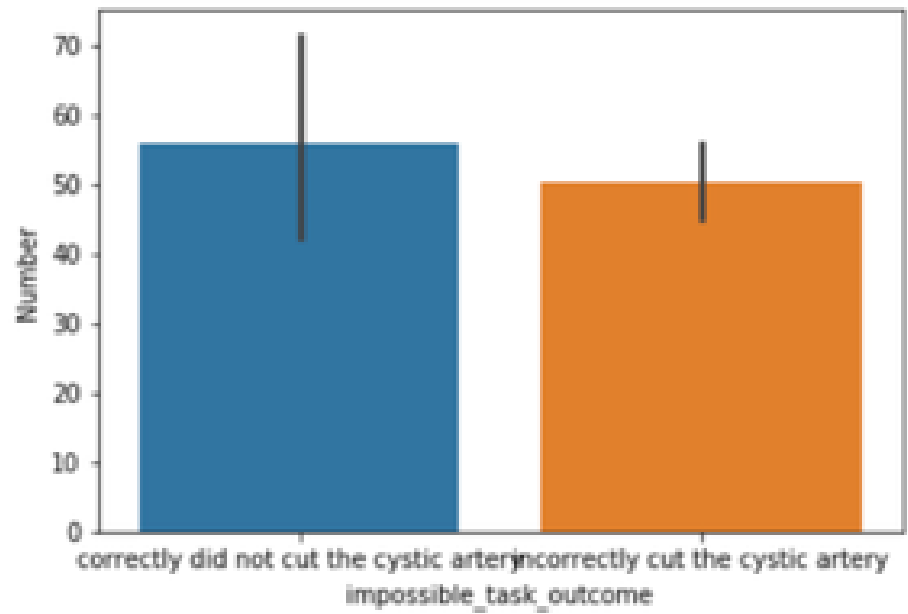
Order –
Feasible: 193.4
Impossible: 169.5



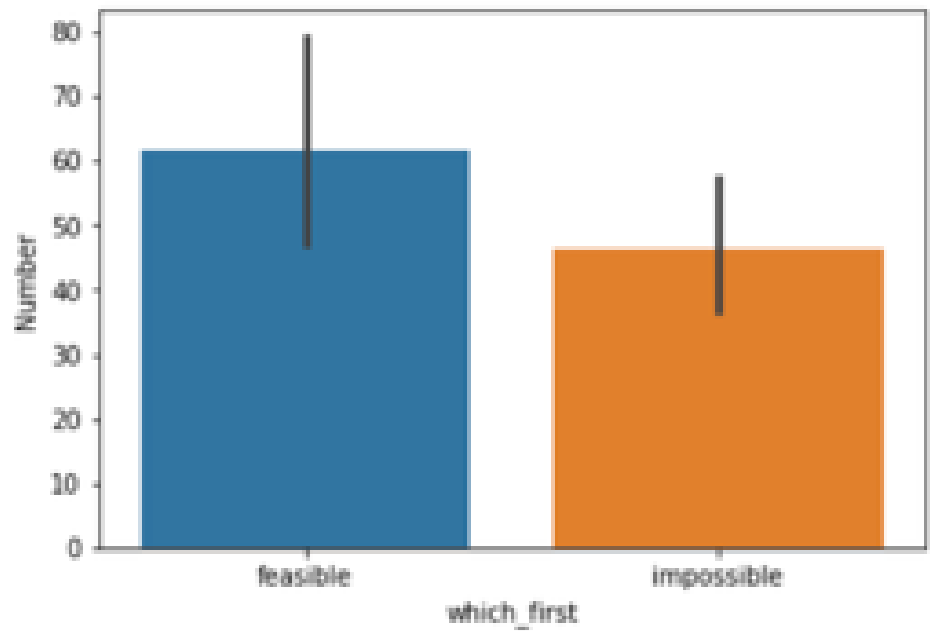
c) Blinks
Difficulty –
Feasible: 48.85
Impossible: 42.15



Outcome –
Feasible: 55.71
Impossible: 50.33



Order –
Feasible: 61.8
Impossible: 46.4



CHAPTER 4

Discussion & Future Work

World-wide, more than 200 million surgeries are performed every year with recent reports revealing that adverse event rates for surgical conditions remain extremely high, despite multiple nationwide and global patient safety initiatives over the past decade[100].

Thus far there has been little work documenting whether there are any physiological markers associated with the making of an error in surgery, however the association between physiological changes and increased cognitive loading has been well documented. Eye-tracking has been used successfully to identify that changes in gaze metrics correlate with increased task difficulty and thus increased cognitive loading. Task-evoked pupil response (TEPR) has been extensively researched and is well proven as a measure of changes in mental work load [112]–[115] Hess and Polt [116] found that there is gradual pupil dilation when a person is preparing to answer a multiplication problem and that dilation reaches a peak immediately before the answer is orally reported, after which it rapidly constricts back to the original size. This would suggest that as the brain performs active problem solving, a physical manifestation of cognitive load can be measured using change in pupil diameter. Across studies, it has been found that mean pupil diameter change is higher during dynamic reasoning tasks which involve planning and high visual attention compared to logical reasoning tasks involving high levels of verbal working memory [117]. Pupil diameter was also shown to correlate highly with error rate during a nuclear power plant simulation in a study conducted by Gao. et al in 2013 [118] There has been some debate about the interference of the light reflex with pupillary dilation. The index of Cognitive Activity (ICA) is a measure that which dissociates the pupil response to cognitive activity from effects of light input. It is a measure of frequency of rapid pupil dilations.[119]

Reduced blink rate and frequency have also been used to identify increased task complexity across multiple fields. The aviation industry has well documented evidence where increased visual demand has been shown to yield a reduced blink rate [120]. It should be noted however that the reduced blink rate may not be constant as in a study conducted by Fairclough et al. reduced blink frequency during episodes of high demand was only found for the first half of a 64 minute task, suggesting that there is an element of adjustment. Although this correlation is well documented, Gao. et al found that following periods of higher cognitive load, a burst of blinks is sometimes observed [118]. This may be because the blinks are delayed until active decision making is complete. Blink rate was also found to be higher preceding incorrect responses than correct responses in a study conducted by Holland and Tarlow [121]

Cardiac activity has been commonly used to measure mental workload and is most often measured via ECG techniques. It is now well known that heart rate (HR) increases with increased task demands, with the introduction of additional memory load, and with having to multi-task [122][123] Sosnowski et al. found that there was a larger increase in HR during tasks requiring problem solving compared to tasks requiring a step-wise completion of a series of elements [124]. Within the surgical field, ambulatory monitoring has been used to measure the HR of surgeons while

they operated. It has been conclusively found that mean HR intra-operatively is much higher compared to pre and post theatre [125] and that the episodes of peak heart rate correlate with increased perceived stress [126][127].

More recently studies have found that skin measurements, more specifically skin conductance, have a role to play in workload measurement and prediction. Electrodermal Activity (EDA) in particular has been shown to be particularly sensitive to sudden changes in external stimuli [128]. Miyake. Et al undertook a computer based task, EDA correlated with task difficulty during a multi attribute task and showed better test-retest reliability than other physiological measures [129] similarly Fairclough and Venables found that skin conductance level increased significantly from baseline decreasing over time during a multi attribute task [130].

The studies presented here are the first of their kind in attempting to use physiological metrics to predict surgeon error intra-operatively. The analysis suggests that skin conductance is the single most relevant metric (of the metrics tested), which can be used to predict the likelihood of a surgeon making an error.

It is known that fluctuations in arousal are regulated by the autonomic nervous system, which is mainly controlled the synergistic activity of the parasympathetic and sympathetic systems. Holper et al also showed that an error memory effect exists, in that participants in their study who recalled making an error at a previous decision point tended to elicit reduced skin conductance responses compared to those who did not remember and repeated the error.[131] My study shows that skin conductance spikes are a strong indicator of impending surgical error. During the course of the analysis, previous error was also found to be one of the most predictive features in the model, and it is important to note that 2/3 of surgeons in the study made repeat error.

The most important skin conductance metrics were a) several measures of autocorrelation and b) the longest strike above the mean. Autocorrelation with a given lag i.e. the correlation of skin conductance with itself, offset by x number of timesteps measured sudden changes; so, if nothing changed, correlation was perfect, as soon as change occurred, correlation dropped. Multiple versions with different lags were tested to avoid being biased by phasing. The “longest strike above the mean” refers to the greatest stretch of time that conductance remained above the mean. I found that baseline conductance fluctuates around mean, resulting in short strikes above mean, but stress induced a high conductance above mean, resulting in a longer strike.

Skin conductance was followed by heart rate and fatigue score. Recognition that an impossible instruction was given, and successfully avoiding error causes greater HR than simply being unaware of one's mistake. Those who finished the short impossible task correctly had an average heart rate of approximately 84 compared to those who finished the short impossible task incorrectly, who had an average HR of 74. Skin conductance standard deviation is almost twice as large (.44 vs .25) in the short impossible task versus the short feasible task and more than 5x bigger in incorrect outcomes (.69) vs correct outcomes (.13) for the short impossible task.

Skin conductance was twice as likely to successfully predict impending error compared to heart rate and fatigue severity score. Gaze metrics were found to be unimportant thus rendering our hypothesis incorrect. However, when I added gaze features, overall model performance improved by 6.4%. This indicates that gaze tracking is important, and future work could focus on identifying this importance further.

There are some critiques of my study, which may have led to the importance of gaze features all scoring less than 0. One of the problems I encountered was the AOI data, having been logged manually. Were I to repeat the experiment, the AOI data may be better featured in a finer resolution than 5 second segments. Another challenge was the limited number of errors. Much of the difference in performance metrics was due to changing how I classify a single or very small numbers of errors. For these experiments to be more robust, I would need more examples of errors. It would also be worth exploring whether simple gaze tracking metrics differed in particular segments, in particular, ocular divergence. Most eye trackers measure the positions of the left and right eyes independently. This allows the extraction of *vergence*, i.e., whether left and right eyes move together or apart from each other, and is labelled as X and Y. This phenomenon is just a natural consequence of focusing near and far. Divergence often happens when our mind drifts away, when losing focus or concentration.

Overall, the focus of this MPhil was the perceptual contributions to decision-making. This project aimed to better understand the psychophysical behaviour associated with making a decision. Decision-making can be regarded as the cognitive process resulting in the selection of a belief or a course of action among several alternative possibilities. When something goes wrong in the decision making process – errors occur. In the context of healthcare, this can have negative health outcomes and potentially lethal consequences.

Although gaze metrics were found to be unimportant in my studies, addition of gaze metrics did lead to an improvement in model performance. I found skin conductance to be the strongest physiological metric in the prediction of a surgeon making an error intra-operatively; as skin conductance increases in relation to increased cognitive load, a surgeon becomes more likely to make an error. This has important implications in the surgical field, as skin conductance could potentially be used as a warning system to indicate when surgeons are at high risk of making an error. The potential for reduction in surgical error rate and improvement in patient safety are important factors to consider.

Appendix

Acronyms

Visual Probe Task: Participants are situated in front of a computer screen and are asked to stare at a fixation cross on the centre of the screen. Two stimuli, one of which is neutral and one of which is being investigated, appear randomly on either side of the screen. The stimuli are presented for a predetermined length of time (most commonly 500ms), before a dot is presented in the location of one former stimulus. Participants are instructed to indicate the location of this dot as quickly as possible, either via keyboard or response box. (https://en.wikipedia.org/wiki/Dot-probe_paradigm)

Antisaccade task: A task which requires participants to execute antisaccades; that is, to look in the opposite direction of a stimulus that automatically catches attention by appearing in the peripheral visual field. Typically participants are required to fixate on a central cross in the middle of a screen as images appear in the periphery. High rates of prosaccades to the peripheral stimuli are considered indicators of increased impulsivity. [132]

Prosaccade: When one looks towards a stimulus

***Gaze direction bias** defined as the proportion of trials on which the first fixation was directed to a food stimulus versus a non-food stimulus (a score above 50% indicates a higher proportion of first fixations on food stimuli) [133]

&**Gaze duration bias** was a measure for **early attention maintenance**, defined as the difference between the sum of fixations between food and non-food stimuli, before gaze was shifted away from the initially fixated picture (a positive score indicates longer initial gaze on food, than non-food, stimuli) [133]

€**Gaze dwell time bias** was made up of the average total dwell time on food versus non-food stimuli (a positive score indicating a bias towards food stimuli) [133]

References

- [1] M. K. Eckstein, B. Guerra-Carrillo, A. T. M. Singley, and S. A. Bunge, "Beyond eye gaze: What else can eyetracking reveal about cognition and cognitive development?," *Dev. Cogn. Neurosci.*, vol. 25, pp. 69–91, 2017.
- [2] K. Holmqvist, M. Nyström, R. Andersson, R. Dewhurst, H. Jarodzka, and J. Van De Weijer, "Eye Tracking: A comprehensive guide to methods and measures." p. 560, 2011.
- [3] B. W. Tatler, N. J. Wade, H. Kwan, J. M. Findlay, and B. M. Velichkovsky, "Yarbus, eye movements, and vision," *Iperception.*, vol. 1, no. 1, pp. 7–27, 2010.
- [4] M. R. Greene, T. Liu, and J. M. Wolfe, "Reconsidering Yarbus: a failure to predict observers' task from eye movement patterns," *Vision Res.*, vol. 62, pp. 1–8, Jun. 2012.
- [5] T. Eggert, "Eye movement recordings: methods," *Dev. Ophthalmol.*, vol. 40, pp. 15–34, 2007.
- [6] K. Colvin, "Is pilots' visual scanning adequate to avoid mid-air collisions," *Proc. ...*, no. Vmc, pp. 1–6, 2005.
- [7] X. Zhang and S.-M. Yuan, *An Eye Tracking Analysis for Video Advertising: Relationship between Advertisement Elements and Effectiveness*, vol. 6. 2018.
- [8] G. Wood and M. R. Wilson, "Quiet-eye training for soccer penalty kicks," *Cogn. Process.*, vol. 12, no. 3, pp. 257–266, 2011.
- [9] I. C. Murray, B. W. Fleck, H. M. Brash, M. E. MacRae, L. L. Tan, and R. a. Minns, "Feasibility of Saccadic Vector Optokinetic Perimetry. A Method of Automated Static Perimetry for Children Using Eye Tracking," *Ophthalmology*, vol. 116, no. 10, pp. 2017–2026, 2009.
- [10] D. Wroblewski, B. a. Francis, A. Sadun, G. Vakili, and V. Chopra, "Testing of Visual Field with Virtual Reality Goggles in Manual and Visual Grasp Modes," *Biomed Res. Int.*, vol. 2014, pp. 1–10, 2014.
- [11] G. Hardiess, E. Papageorgiou, U. Schiefer, and H. a. Mallot, "Functional compensation of visual field deficits in hemianopic patients under the influence of different task demands," *Vision Res.*, vol. 50, no. 12, pp. 1158–1172, 2010.
- [12] P. a. Aspinall, S. Borooah, C. Al Alouch, J. Roe, a Laude, R. Gupta, M. Gupta, a Montarzino, and B. Dhillon, "Gaze and pupil changes during navigation in age-related macular degeneration.," *Br. J. Ophthalmol.*, pp. 1393–1397, 2014.
- [13] D. P. Crabb, N. D. Smith, and H. Zhu, "What's on TV? Detecting age-related neurodegenerative eye disease using eye movement scanpaths," *Front. Aging Neurosci.*, vol. 6, no. November, pp. 1–10, 2014.
- [14] N. D. Smith, F. C. Glen, V. M. Mönter, and D. P. Crabb, "Using eye tracking to assess reading performance in patients with glaucoma: A within-person study," *J. Ophthalmol.*, vol. 2014, 2014.
- [15] P. R. Jones, S. Kalwarowsky, J. Atkinson, O. J. Braddick, and M. Nardini, "Automated Measurement of Resolution Acuity in Infants Using Remote Eye-Tracking," *Invest. Ophthalmol. Vis. Sci.*, vol. 55, no. 12, pp. 8102–8110, 2014.
- [16] E. Ahtola, S. Stjerna, S. Yrttiaho, C. a. Nelson, J. M. Leppänen, and S. Vanhatalo, "Dynamic eye tracking based metrics for infant gaze patterns in the face-distractor competition paradigm," *PLoS One*, vol. 9, no. 5, pp. 1–11, 2014.
- [17] J. L. Richmond, J. L. Zhao, and M. a. Burns, "What goes where? Eye tracking reveals spatial relational memory during infancy," *J. Exp. Child Psychol.*, vol. 130, pp. 79–91, 2015.
- [18] A. Hammar, A. Lund, and K. Hugdahl, "Selective impairment in effortful information processing in major depression.," *J. Int. Neuropsychol. Soc.*, vol. 9, no. 6, pp. 954–959, 2003.
- [19] J. L. Kellough, C. G. Beevers, A. J. Ellis, and T. T. Wells, "Time Course of Selective Attention in Clinically Depressed Young Adults: An Eye Tracking Study," vol. 46, no.

- 11, pp. 1238–1243, 2009.
- [20] L. Isaac, J. N. Vrijssen, M. Rinck, A. Speckens, and E. S. Becker, "Shorter gaze duration for happy faces in current but not remitted depression: Evidence from eye movements," *Psychiatry Res.*, vol. 218, no. 1–2, pp. 79–86, 2014.
- [21] C. G. Beevers, H. J. Lee, T. T. Wells, A. J. Ellis, and M. J. Telch, "Association of predeployment gaze bias for emotion stimuli with later symptoms of PTSD and depression in soldiers deployed in Iraq," *Am. J. Psychiatry*, vol. 168, no. 7, pp. 735–741, 2011.
- [22] J. W. Weeks, A. N. Howell, and P. R. Goldin, "Gaze avoidance in social anxiety disorder," *Depress. Anxiety*, vol. 30, no. 8, pp. 749–756, 2013.
- [23] S. Chen, R. Zhou, H. Cui, and X. Chen, "Deficits in cue detection underlie event-based prospective memory impairment in major depression: An eye tracking study," *Psychiatry Res.*, vol. 209, no. 3, pp. 453–458, 2013.
- [24] C. Liossi, D. E. Schoth, H. J. Godwin, and S. P. Liversedge, "Author ' s personal copy Using eye movements to investigate selective attention in chronic daily headache," *Pain*, vol. 155, pp. 503–510, 2014.
- [25] D. M. Isaacowitz, K. Toner, D. Goren, and H. R. Wilson, "Positive Gaze in Older Adults," *October*, vol. 19, no. 9, pp. 848–853, 2009.
- [26] J. Katsanis, J. Taylor, W. G. Iacono, and M. a Hammer, "Heritability of different measures of smooth pursuit eye tracking dysfunction: a study of normal twins.," *Psychophysiology*, vol. 37, no. 6, pp. 724–730, 2000.
- [27] W. M. Clementz, B.A., Iacono, W.G., Grove, "The construct validity of root-mean-square error for quantifying smooth-pursuit eye tracking abnormalities in schizophrenia." *Biological Psychiatry*.
- [28] D. L. Levy, A. B. Sereno, D. C. Gooding, and G. a. O'Driscoll, "Eye tracking dysfunction in schizophrenia: Characterization and pathophysiology," *Curr. Top. Behav. Neurosci.*, vol. 2010, no. 4, pp. 311–347, 2010.
- [29] P. S. Holzman, D. L. Levy, S. Matthysse, and N. R. Mendell, "Eye tracking dysfunction and schizophrenia: a critical perspective.," *Schizophrenia bulletin*, vol. 19, no. 3. pp. 461–536, 1993.
- [30] R. G. Ross, A. Olincy, J. G. Harris, A. Radant, L. E. Adler, and R. Freedman, "Anticipatory saccades during smooth pursuit eye movements and familial transmission of schizophrenia," *Biol. Psychiatry*, vol. 44, no. 8, pp. 690–697, 1998.
- [31] R. Lencer, K. Trillenber-Krecker, E. Schwinger, and V. Arolt, "Schizophrenia spectrum disorders and eye tracking dysfunction in singleton and multiplex schizophrenia families," *Schizophr. Res.*, vol. 60, no. 1, pp. 33–45, 2003.
- [32] A. Sporn, D. Greenstein, N. Gogtay, F. Sailer, D. W. Hommer, R. Rawlings, R. Nicolson, M. F. Egan, M. Lenane, P. Gochman, D. R. Weinberger, and J. L. Rapoport, "Childhood-onset schizophrenia: Smooth pursuit eye-tracking dysfunction in family members," *Schizophr. Res.*, vol. 73, no. 2–3, pp. 243–252, 2005.
- [33] T. L. Hodgson, B. Tiesman, A. M. Owen, and C. Kennard, "Abnormal gaze strategies during problem solving in Parkinson ' s disease," vol. 40, pp. 411–422, 2002.
- [34] T. J. Crawford, A. Devereaux, S. Higham, and C. Kelly, "The disengagement of visual attention in Alzheimer ' s disease : a longitudinal eye-tracking study," vol. 7, no. June, pp. 1–10, 2015.
- [35] J. Keller, M. Gorges, H. T. Horn, H. E. a. Aho-Özhan, E. H. Pinkhardt, I. Uttner, J. Kassubek, A. C. Ludolph, and D. Lulé, "Eye-tracking controlled cognitive function tests in patients with amyotrophic lateral sclerosis: a controlled proof-of-principle study," *J. Neurol.*, 2015.
- [36] S. S. Hall, A. M. Y. A. Lightbody, and L. C. Huffman, "Physiological Correlates of Social Avoidance Behavior in Children and Adolescents With Fragile X Syndrome," *J. Am. Acad. Child Adolesc. Psychiatry*, vol. 48, no. 3, pp. 320–329, 2009.
- [37] and R. K. Jinger Pana, Ming Yana, Jochen Laubrocka, Hua Shub, "Saccade-target selection of dyslexic children when reading Chinese," *Vision Res.*, pp. 24–30, 2015.
- [38] S. a. Rose, A. Djukic, J. J. Jankowski, J. F. Feldman, I. Fishman, and M. Valicenti-

- Mcdermott, "Rett syndrome: An eye-tracking study of attention and recognition memory," *Dev. Med. Child Neurol.*, vol. 55, no. 4, pp. 364–371, 2013.
- [39] K. Chawarska, S. MacAri, and F. Shic, "Context modulates attention to social scenes in toddlers with autism," *J. Child Psychol. Psychiatry Allied Discip.*, vol. 53, no. 8, pp. 903–913, 2012.
- [40] T. C. J. de Wit, T. Falck-Ytter, and C. von Hofsten, "Young children with Autism Spectrum Disorder look differently at positive versus negative emotional faces," *Res. Autism Spectr. Disord.*, vol. 2, no. 4, pp. 651–659, 2008.
- [41] M. Hosozawa, K. Tanaka, T. Shimizu, T. Nakano, and S. Kitazawa, "How Children With Specific Language Impairment View Social Situations: An Eye Tracking Study," *Pediatrics*, vol. 129, no. 6, pp. e1453–e1460, 2012.
- [42] W. Jones, K. Carr, and A. Klin, "Absence of preferential looking to the eyes of approaching adults predicts level of social disability in 2-year-old toddlers with autism spectrum disorder.," *Arch. Gen. Psychiatry*, vol. 65, no. 8, pp. 946–954, 2008.
- [43] N. Merin, G. S. Young, S. Ozonoff, and S. J. Rogers, "Visual fixation patterns during reciprocal social interaction distinguish a subgroup of 6-month-old infants at-risk for autism from comparison infants," *J. Autism Dev. Disord.*, vol. 37, no. 1, pp. 108–121, 2007.
- [44] T. Nakano, K. Tanaka, Y. Endo, Y. Yamane, T. Yamamoto, Y. Nakano, H. Ohta, N. Kato, and S. Kitazawa, "Atypical gaze patterns in children and adults with autism spectrum disorders dissociated from developmental changes in gaze behaviour.," *Proc. Biol. Sci.*, vol. 277, no. 1696, pp. 2935–2943, 2010.
- [45] K. Rice, J. M. Moriuchi, W. Jones, and A. Klin, "Parsing heterogeneity in autism spectrum disorders: Visual scanning of dynamic social scenes in school-aged children," *J. Am. Acad. Child Adolesc. Psychiatry*, vol. 51, no. 3, pp. 238–248, 2012.
- [46] A. S. a Chetwood, K. W. Kwok, L. W. Sun, G. P. Mylonas, J. Clark, A. Darzi, and G. Z. Yang, "Collaborative eye tracking: A potential training tool in laparoscopic surgery," *Surg. Endosc. Other Interv. Tech.*, vol. 26, no. 7, pp. 2003–2009, 2012.
- [47] G. Wood, J. Batt, a. Appelboom, a. Harris, and M. R. Wilson, "Exploring the Impact of Expertise, Clinical History, and Visual Search on Electrocardiogram Interpretation," *Med. Decis. Mak.*, vol. 34, no. 1, pp. 75–83, 2013.
- [48] R. R. Bond, T. Zhu, D. D. Finlay, B. Drew, P. D. Kligfield, D. Guldenring, C. Breen, a. G. Gallagher, M. J. Daly, and G. D. Clifford, "Assessing computerized eye tracking technology for gaining insight into expert interpretation of the 12-lead electrocardiogram: An objective quantitative approach," *J. Electrocardiol.*, vol. 47, no. 6, pp. 895–906, 2014.
- [49] M. R. Wilson, S. J. Vine, E. Bright, R. S. W. Masters, D. Defriend, and J. S. McGrath, "Gaze training enhances laparoscopic technical skill acquisition and multi-tasking performance: A randomized, controlled study," *Surg. Endosc. Other Interv. Tech.*, vol. 25, no. 12, pp. 3731–3739, 2011.
- [50] S. J. Vine, R. S. W. Masters, J. S. McGrath, E. Bright, and M. R. Wilson, "Cheating experience: Guiding novices to adopt the gaze strategies of experts expedites the learning of technical laparoscopic skills," *Surg. (United States)*, vol. 152, no. 1, pp. 32–40, 2012.
- [51] J. Causer, J. N. Vickers, R. Snelgrove, G. Arsenault, and A. Harvey, "Performing under pressure: Quiet eye training improves surgical knot-tying performance," *Surgery*, vol. 156, no. 5, pp. 1089–1096, 2014.
- [52] P. Thiagarajan, K. J. Ciuffreda, J. E. Capo-Aponte, D. P. Ludlam, and N. Kapoor, "Oculomotor neurorehabilitation for reading in mild traumatic brain injury (mTBI): An integrative approach," *NeuroRehabilitation*, vol. 34, no. 1, pp. 129–146, 2014.
- [53] N. Ahmidi, G. D. Hager, L. Ishii, G. Fichtinger, G. L. Gallia, and M. Ishii, "Surgical gesture classification from eye tracking and tool motion in Minimally Invasive Surgery," *Lect. notes Comput. Sci.*, vol. 63, no. 63, pp. 295–302, 2010.
- [54] P. O'Meara, G. Munro, B. Williams, S. Cooper, F. Bogossian, L. Ross, L. Sparkes, M. Browning, and M. McClounan, "Developing situation awareness amongst nursing and

- paramedicine students utilizing eye tracking technology and video debriefing techniques: A proof of concept paper," *Int. Emerg. Nurs.*, vol. 23, no. 2, pp. 94–99, 2015.
- [55] J. L. Marquard, P. L. Henneman, Z. He, J. Jo, D. L. Fisher, and E. a. Henneman, "Nurses' behaviors and visual scanning patterns may reduce patient identification errors.," *J. Exp. Psychol. Appl.*, vol. 17, no. 3, pp. 247–256, 2011.
 - [56] A. Szulewski and D. Howes, "Combining first-person video and gaze-tracking in medical simulation: A technical feasibility study," *Sci. World J.*, vol. 2014, 2014.
 - [57] E. a. Henneman, H. Cunningham, D. L. Fisher, K. Plotkin, B. H. Nathanson, J. P. Roche, J. L. Marquard, C. a. Reilly, and P. L. Henneman, "Eye tracking as a debriefing mechanism in the simulated setting improves patient safety practices.," *Dimens. Crit. Care Nurs.*, vol. 33, no. 3, pp. 129–35, 2014.
 - [58] R. S. a Khan, G. Tien, M. S. Atkins, B. Zheng, O. N. M. Panton, and A. T. Meneghetti, "Analysis of eye gaze: Do novice surgeons look at the same location as expert surgeons during a laparoscopic operation?," *Surg. Endosc. Other Interv. Tech.*, vol. 26, no. 12, pp. 3536–3540, 2012.
 - [59] A. Ghebremichael, "Gaze patterns of anatomy students through classroom learning and familiarization," pp. 2014–2015, 2014.
 - [60] B. P. Soh, W. M. Reed, A. Poulos, and P. C. Brennan, "E-tutorial improves students' ability to detect lesions.," *Radiol. Technol.*, vol. 85, no. 1, pp. 17–26, 2013.
 - [61] M. H. Sodergren, F. Orihuela-Espina, F. Froghi, J. Clark, J. Teare, G. Z. Yang, and a. Darzi, "Value of orientation training in laparoscopic cholecystectomy," *Br. J. Surg.*, vol. 98, no. 10, pp. 1437–1445, 2011.
 - [62] M. Wilson, J. McGrath, S. Vine, J. Brewer, D. Defriend, and R. Masters, "Psychomotor control in a virtual laparoscopic surgery training environment: Gaze control parameters differentiate novices from experts," *Surg. Endosc. Other Interv. Tech.*, vol. 24, no. 10, pp. 2458–2464, 2010.
 - [63] A. Harvey, J. N. Vickers, R. Snelgrove, M. F. Scott, and S. Morrison, "Expert surgeon's quiet eye and slowing down: Expertise differences in performance and quiet eye duration during identification and dissection of the recurrent laryngeal nerve," *Am. J. Surg.*, vol. 207, no. 2, pp. 187–193, 2014.
 - [64] L. Richstone, M. J. Schwartz, C. Seideman, J. Cadeddu, S. Marshall, and L. R. Kavoussi, "Eye metrics as an objective assessment of surgical skill.," *Ann. Surg.*, vol. 252, no. 1, pp. 177–182, 2010.
 - [65] T. Tien, P. H. Pucher, M. H. Sodergren, K. Sriskandarajah, G. Z. Yang, and A. Darzi, "Differences in gaze behaviour of expert and junior surgeons performing open inguinal hernia repair," *Surg. Endosc.*, pp. 405–413, 2014.
 - [66] E. S. M. Tiersma, a a W. Peters, H. a Mooij, and G. J. Fleuren, "Visualising scanning patterns of pathologists in the grading of cervical intraepithelial neoplasia.," *J. Clin. Pathol.*, vol. 56, no. 9, pp. 677–680, 2003.
 - [67] E. a. Krupinski, A. a. Tillack, L. Richter, J. T. Henderson, A. K. Bhattacharyya, K. M. Scott, A. R. Graham, M. R. Descour, J. R. Davis, and R. S. Weinstein, "Eye-movement study and human performance using telepathology virtual slides. Implications for medical education and differences with experience," *Hum. Pathol.*, vol. 37, no. 12, pp. 1543–1556, 2006.
 - [68] H. L. Kundel, C. F. Nodine, E. a. Krupinski, and C. Mello-Thoms, "Using Gaze-tracking Data and Mixture Distribution Analysis to Support a Holistic Model for the Detection of Cancers on Mammograms," *Acad. Radiol.*, vol. 15, no. 7, pp. 881–886, 2008.
 - [69] B. Anderson and C.-R. Shyu, "A preliminary study to understand tacit knowledge and visual routines of medical experts through gaze tracking.," *AMIA Annu. Symp. Proc.*, vol. 2010, pp. 21–25, 2010.
 - [70] H. L. Kundel, C. F. Nodine, and L. Toto, "Searching for lung nodules. The guidance of visual scanning.," *Investigative radiology*, vol. 26, no. 9, pp. 777–781, 1991.
 - [71] G. D. Rubin, J. K. Lyo, D. S. Paik, A. J. Sherbondy, L. C. Chow, A. N. Leung, R.

- Mindelzun, P. K. Schraedley-Desmond, S. E. Zinck, D. P. Naidich, and S. Napel, "Pulmonary nodules on multi-detector row CT scans: performance comparison of radiologists and computer-aided detection.," *Radiology*, vol. 234, no. 1, pp. 274–283, 2005.
- [72] P. F. Pinsky, D. S. Gierada, P. H. Nath, E. Kazerooni, and J. Amorosa, "National lung screening trial: variability in nodule detection rates in chest CT studies.," *Radiology*, vol. 268, no. 3, pp. 865–73, 2013.
- [73] J. E. Roos, D. Paik, D. Olsen, E. G. Liu, L. C. Chow, A. N. Leung, R. Mindelzun, K. R. Choudhury, D. P. Naidich, S. Napel, and G. D. Rubin, "Computer-aided detection (CAD) of lung nodules in CT scans: Radiologist performance and reading time with incremental CAD assistance," *Eur. Radiol.*, vol. 20, no. 3, pp. 549–557, 2010.
- [74] G. D. Rubin, J. E. Roos, B. Harrawood, D. L. Ly, D. M. Seaman, and L. M. Hurwitz, "Recognition , and Decision in the Detection of Lung Nodules on CT Scans : Elucidation with Eye Tracking 1," *Radiology*, vol. 274, no. 1, 2015.
- [75] A. Meining, S. Atasoy, A. Chung, N. Navab, and G. Z. Yang, "'Eye-tracking' for assessment of image perception in gastrointestinal endoscopy with narrow-band imaging compared with white-light endoscopy," *Endoscopy*, vol. 42, no. 8, pp. 652–655, 2010.
- [76] C. Almansa, M. W. Shahid, M. G. Heckman, S. Preissler, and M. B. Wallace, "Association between visual gaze patterns and adenoma detection rate during colonoscopy: a preliminary investigation.," *Am. J. Gastroenterol.*, vol. 106, no. 6, pp. 1070–1074, 2011.
- [77] P. Phillips, D. Boone, S. Mallett, S. A. Taylor, M. B. Bs, D. G. Altman, D. Manning, A. Gale, S. Halligan, and M. B. Bs, "Method for Tracking Eye Gaze during Interpretation of Endoluminal 3D CT Colonography : Technical Description," vol. 267, no. 3, 2013.
- [78] T. Jaarsma, H. Jarodzka, M. Nap, J. J. G. Van Merrienboer, and H. P. a Boshuizen, "Expertise under the microscope: Processing histopathological slides," *Med. Educ.*, vol. 48, no. 3, pp. 292–300, 2014.
- [79] G. Wood, K. M. Knapp, B. Rock, C. Cousens, C. Roobottom, and M. R. Wilson, "Visual expertise in detecting and diagnosing skeletal fractures," *Skeletal Radiol.*, vol. 42, no. 2, pp. 165–172, 2013.
- [80] E. C. O'Neill, Y. X. G. Kong, P. P. Connell, D. N. Ong, S. a. Haymes, M. a. Coote, and J. G. Crowston, "Gaze behavior among experts and trainees during optic disc examination: Does how we look affect what we see?," *Investig. Ophthalmol. Vis. Sci.*, vol. 52, no. 7, pp. 3976–3983, 2011.
- [81] G. Tourassi, S. Voisin, V. Paquit, and E. Krupinski, "Investigating the link between radiologists' gaze, diagnostic decision, and image content.," *J. Am. Med. Inform. Assoc.*, vol. 20, no. 6, pp. 1067–75, 2013.
- [82] J. S. Baschnagel, "Using mobile eye-tracking to assess attention to smoking cues in a naturalized environment," *Addict. Behav.*, vol. 38, no. 12, pp. 2837–2840, 2013.
- [83] O. M. Maynard, A. Attwood, L. O'Brien, S. Brooks, C. Hedge, U. Leonards, and M. R. Munafò, "Avoidance of cigarette pack health warnings among regular cigarette smokers," *Drug Alcohol Depend.*, vol. 136, no. 1, pp. 170–174, 2014.
- [84] M. Shankleman, C. Sykes, K. L. Mandeville, S. Di Costa, and K. Yarrow, "Standardised (plain) cigarette packaging increases attention to both text-based and graphical health warnings: experimental evidence," *Public Health*, vol. 129, no. 1, pp. 37–42, 2015.
- [85] D. G. Gilbert, N. E. Rabinovich, D. Malpass, J. Mrnak, H. Riise, L. Adams, C. Sugai, and M. DevleschHoward, "Effects of nicotine on affect are moderated by stressor proximity and frequency, positive alternatives, and smoker status.," *Nicotine Tob. Res.*, vol. 10, no. 7, pp. 1171–1183, 2008.
- [86] N. J. Wachter, "Nicotine Differentially Modulates Antisaccade Eye-Gaze Away from Emotional Stimuli in Nonsmokers Stratified by Pre-Task Baseline Performance," *Pyschopharmacology (Berlin)*, vol. 29, no. 6, pp. 997–1003, 2012.
- [87] S. Cairney, P. Maruff, C. B. Burns, J. Currie, and B. J. Currie, "Saccade dysfunction

- associated with chronic petrol sniffing and lead encephalopathy.," *J. Neurol. Neurosurg. Psychiatry*, vol. 75, no. 3, pp. 472–476, 2004.
- [88] A. K. Rose, K. Brown, M. Field, and L. Hogarth, "The contributions of value-based decision-making and attentional bias to alcohol-seeking following devaluation," *Addiction*, vol. 108, no. 7, pp. 1241–1249, 2013.
 - [89] P. J. Brown, J. L. Marquard, B. Amster, M. Romoser, J. Friderici, S. Goff, and D. Fisher, "What Do Physicians Read (and Ignore) in Electronic Progress Notes?," *Appl. Clin. Inform.*, vol. 5, no. 2, pp. 430–444, 2014.
 - [90] E. Sutton, Y. Youssef, N. Meenaghan, C. Godinez, Y. Xiao, T. Lee, D. Dexter, and A. Park, "Gaze disruptions experienced by the laparoscopic operating surgeon," *Surg. Endosc. Other Interv. Tech.*, vol. 24, no. 6, pp. 1240–1244, 2010.
 - [91] and B. B. E Montague, J Xu, O Asan, P Chen, B Chewning, "Modeling Eye Gaze Patterns in Clinician-Patient Interaction with Lag Sequential Analysis," *Hum. Factors*, vol. 53, no. 5, pp. 502–516, 2011.
 - [92] O. Asan, J. Xu, and E. Montague, "Dynamic Comparison of Physicians' Interaction Style with Electronic Health Records in Primary Care Settings," *J. Gen. Pract.*, vol. 29, no. 6, pp. 997–1003, 2012.
 - [93] M. H. Sodergren, F. Orihuela-Espina, J. Clark, A. Darzi, and G. Z. Yang, "A hidden Markov model-based analysis framework using eye-tracking data to characterise re-orientation strategies in minimally invasive surgery," *Cogn. Process.*, vol. 11, no. 3, pp. 275–283, 2010.
 - [94] M. H. Sodergren, F. Orihuela-Espina, P. Mountney, J. Clark, J. Teare, A. Darzi, and G.-Z. Yang, "Orientation strategies in natural orifice transluminal endoscopic surgery.," *Ann. Surg.*, vol. 254, no. 2, pp. 257–266, 2011.
 - [95] D. P. Noonan, G. P. Mylonas, a. Darzi, and G.-Z. Y. G.-Z. Yang, "Gaze contingent articulated robot control for robot assisted minimally invasive surgery," *2008 IEEE/RSJ Int. Conf. Intell. Robot. Syst.*, pp. 22–26, 2008.
 - [96] U. Lahiri, E. Bekele, E. Dohrmann, Z. Warren, and N. Sarkar, "Design of a Virtual Reality Based Adaptive Response Technology for Children With Autism," *Trans Neural Syst Rehabil Eng.*, vol. 21, no. 1, pp. 1–27, 2013.
 - [97] F. Maringelli, N. Brienza, F. Scorrano, F. Grasso, and C. Gregoretti, "Gaze-controlled, computer-assisted communication in Intensive Care Unit: 'speaking through the eyes'.," *Minerva Anesthesiol.*, vol. 79, no. 2, pp. 165–75, 2013.
 - [98] J. J. van Middendorp, F. Watkins, C. Park, and H. Landymore, "Eye-tracking computer systems for inpatients with tetraplegia: findings from a feasibility study," *Spinal Cord*, vol. 53, no. 3, pp. 221–225, 2014.
 - [99] R. Spataro, M. Ciriaco, C. Manno, and V. La Bella, "The eye-tracking computer device for communication in amyotrophic lateral sclerosis," *Acta Neurol. Scand.*, vol. 130, no. 1, pp. 40–45, 2014.
 - [100] F. J. Kim, R. Donalisio Da Silva, D. Gustafson, L. Nogueira, T. Harlin, and D. L. Paul, "Current issues in patient safety in surgery: a review," *Patient Saf. Surg.*, pp. 2–9, 2015.
 - [101] W. A. Klei, R. G. Hoff, E. E. H. L. Aarnhem, R. K. J. Simmermacher, L. P. E. Regli, T. H. Kappen, L. Wolfswinkel, C. J. Kalkman, W. F. Buhre, and L. M. Peelen, "Effects of the introduction of the WHO 'Surgical Safety Checklist' on in-hospital mortality: a cohort study," *Ann. Surg.*, vol. 255, no. 1, pp. 44–49, 2012.
 - [102] R. Hoffer, TP; Kerr, EA; Hayward, "What is an Error?," *Eff. Clin. Pract.*, vol. 3, no. October, pp. 12–14, 2011.
 - [103] S. van Gaal, F. P. de Lange, and M. X. Cohen, "The role of consciousness in cognitive control and decision making," *Front. Hum. Neurosci.*, vol. 6, no. May, p. 121, 2012.
 - [104] S. Nieuwenhuis, K. R. Ridderinkhof, J. Blom, G. P. Band, and a Kok, "Error-related brain potentials are differentially related to awareness of response errors: evidence from an antisaccade task.," *Psychophysiology*, vol. 38, no. 5, pp. 752–760, 2001.
 - [105] G. Dutilh, D. Van Ravenzwaaij, S. Nieuwenhuis, H. L. J. Van der Maas, B. U.

- Forstmann, and E. J. Wagenmakers, "How to measure post-error slowing: A confound and a simple solution," *J. Math. Psychol.*, vol. 56, no. 3, pp. 208–216, 2012.
- [106] W. Boucsein, "Electrodermal Activity," *Behav. Psychophysiol. Med. Ienum Press.*, 1992.
- [107] A. M. Dawson, M.E. & Schell, "Principles of Psychophysiology: Physical, social, and inferential elements," p. Chapter 8-The Electrodermal System, 1990.
- [108] P. O. Valko, C. L. Bassetti, K. E. Bloch, U. Held, and C. R. Baumann, "Fatigue Severity Scale (FSS, English version)*," *Sleep*, vol. 31, 2008.
- [109] M. L. Cassin, Barbara Rubin, *Dictionary Of Eye Terminology*, 6th ed. Triad Pub Co; 2nd edition (Mar. 1990), 1990.
- [110] W. Koehrsen, "Beyond Accuracy: Precision and Recall," *Towards Data Science*. [Online]. Available: <https://towardsdatascience.com/beyond-accuracy-precision-and-recall-3da06bea9f6c>. [Accessed: 21-Dec-2018].
- [111] Y. Sasaki, "The truth of the F-measure," pp. 1–5, 2007.
- [112] X. Jiang, B. Zheng, G. Tien, and M. S. Atkins, "Pupil Response to Precision in Surgical Task Evaluation_2013.pdf," pp. 7–11.
- [113] R. L. Collins, "Pupil Size in Relation to mental workload in simple problem solving," 1952.
- [114] S. P. Marshall, "Identifying Cognitive State from Eye Metrics," vol. 78, no. 5.
- [115] S. J. Granholm E, Arsarnow R, "Pupillary responses index cognitive resource limitations - Granholm et al 1996.pdf," pp. 457–461.
- [116] E. H. Hess and J. M. Polt, "Pupil Size in Relation to Mental Activity during Simple Problem-Solving," *Science (80-.)*, vol. 143, no. 3611, p. 1190 LP-1192, Mar. 1964.
- [117] P. J. Causse M, Sénard J, Démonet J, "Monitoring Cognitive and Emotional Processes Through Pupil and Cardiac Response During Dynamic Versus Logical Task," *Appl. Psychophysiol. Biofeedback*, vol. 35, no. 2, pp. 115–123, 2010.
- [118] Q. Gao, Y. Wang, F. Song, Z. Li, and X. Dong, "Mental workload measurement for emergency operating procedures in digital nuclear power plants," *Ergonomics*, vol. 56, no. 7, pp. 1070–1085, Jul. 2013.
- [119] J. Vogels, V. Demberg, and J. Kray, "The Index of Cognitive Activity as a Measure of Cognitive Processing Load in Dual Task Settings," *Front. Psychol.*, vol. 9, p. 2276, 2018.
- [120] J. A. Veltman and A. W. K. Gaillard, "Physiological indices of workload in a simulated flight task," *Biol. Psychol.*, vol. 42, no. 3, pp. 323–342, 1996.
- [121] M. K. Holland and G. Tarlow, "Blinking and Mental Load," *Psychol. Rep.*, vol. 31, no. 1, pp. 119–127, 1972.
- [122] L. Finsen, K. Søgaaard, C. Jensen, V. Borg, and H. Christensen, "Muscle activity and cardiovascular response during computer-mouse work with and without memory demands," *Ergonomics*, vol. 44, no. 14, pp. 1312–1329, 2001.
- [123] L. R. Fournier, G. F. Wilson, and C. R. Swain, "Electrophysiological, behavioral, and subjective indexes of workload when performing multiple tasks: manipulations of task difficulty and training," *Int. J. Psychophysiol.*, vol. 31, no. 2, pp. 129–145, 1999.
- [124] T. Sosnowski, B. Krzywosz-Rynkiewicz, and J. Roguska, "Program running versus problem solving: Mental task effect on tonic heart rate," *Psychophysiology*, vol. 41, no. 3, pp. 467–475, 2004.
- [125] G. E. Foster, D. F. Evans, and J. D. Hardcastle, "HEART-RATES OF SURGEONS DURING OPERATIONS AND OTHER CLINICAL ACTIVITIES AND THEIR MODIFICATION BY OXPRENOLOL," *Lancet*, vol. 311, no. 8078, pp. 1323–1325, Jun. 1978.
- [126] K. I. Jones, F. Amawi, A. Bhalla, O. Peacock, J. P. Williams, and J. N. Lund, "Assessing surgeon stress when operating using heart rate variability and the State Trait Anxiety Inventory: will surgery be the death of us?," *Color. Dis.*, vol. 17, no. 4, pp. 335–341, 2015.
- [127] A. Rieger, R. Stoll, S. Kreuzfeld, K. Behrens, and M. Weippert, "Heart rate and heart rate variability as indirect markers of surgeons' intraoperative stress," *Int. Arch.*

- Occup. Environ. Health*, vol. 87, no. 2, pp. 165–174, Feb. 2014.
- [128] C. Collet, E. Salvia, and C. Petit-Boulanger, “Measuring workload with electrodermal activity during common braking actions,” *Ergonomics*, vol. 57, no. 6, pp. 886–896, 2014.
 - [129] S. Miyake, S. Yamada, T. Shoji, Y. Takae, N. Kuge, and T. Yamamura, “Physiological responses to workload change. A test/retest examination,” *Appl. Ergon.*, vol. 40, no. 6, pp. 987–996, 2009.
 - [130] S. H. Fairclough and L. Venables, “Prediction of subjective states from psychophysiology: A multivariate approach,” *Biol. Psychol.*, vol. 71, no. 1, pp. 100–110, 2006.
 - [131] L. Holper, N. Jäger, F. Scholkmann, and M. Wolf, “Error detection and error memory in spatial navigation as reflected by electrodermal activity,” *Cogn. Process.*, vol. 14, no. 4, pp. 377–389, Nov. 2013.
 - [132] K. E. Giel, K. Schag, C. Plewnia, and S. Zipfel, “Antisaccadic training to improve impulsivity in binge eating disorder,” *Eur. Eat. Disord. Rev.*, vol. 21, no. 6, pp. 488–492, 2013.
 - [133] J. Werthmann, A. Roefs, C. Nederkoorn, K. Mogg, B. P. Bradley, and A. Jansen, “Attention bias for food is independent of restraint in healthy weight individuals-An eye tracking study,” *Eat. Behav.*, vol. 14, no. 3, pp. 397–400, 2013.