Author:
Muller-Glodde, Maren

Title:
Emotion recognition after traumatic brain injury

what can the Bristol Emotion Recognition Task tell us?

Emotion Recognition after Traumatic Brain Injury: What can the Bristol Emotion Recognition Task tell us?

Maren Müller-Glodde

June 2022

School of Psychological Science

A dissertation submitted to the University of Bristol in accordance with the requirements for award of the degree of Doctor of Philosophy in the Faculty of Life Sciences

Word Count: 62827

Abstract

Traumatic Brain Injury (TBI) occurs when a head injury results in altered brain function. There is substantial evidence that moderate to severe TBI is associated with decreased emotion recognition, and performance on emotion recognition tasks could be predictive of social functioning after TBI. Despite this, emotion recognition is not routinely assessed in clinical settings because tasks used to measure emotion recognition lack information about psychometrics and are rarely developed for use in a TBI population.

This thesis investigated the psychometric properties of the Bristol Emotion Recognition Task (ERT) by considering measurement reliability and validity of inferences about emotion recognition. Findings suggested low test-retest reliability in a longitudinal cohort, but adequate internal consistency in both neurologically healthy and TBI populations. A construct validation study assessing correlations between the Bristol ERT and other tasks indicated that the Bristol ERT can be used as a measure of emotion recognition. A cross-sectional study was conducted to assess emotion recognition after moderate to severe TBI. Results suggested that participants with TBI performed worse than controls on the Bristol ERT. This difference was not explained by anxiety or other potentially confounding factors. The same study design was used to investigate emotion recognition after mild TBI and found no evidence for a difference in performance between people with mild TBI and controls. Finally, associations between emotion recognition, mild TBI, and anxiety were investigated in a longitudinal cohort. There was no evidence that mild TBI or Generalised Anxiety Disorder were associated with emotion recognition, but weak evidence that participants with mild TBI were more likely to label faces as angry.

The findings highlight the importance of considering severity of injury when assessing emotion recognition after TBI. They also suggest the Bristol ERT has potential as a measure of emotion recognition in clinical settings, but further development is needed.

Dedication and Acknowledgements

First and foremost, I would like to thank my supervisors Professor Ian Penton-Voak and Dr. Robyn Wootton. I am so grateful for their endless encouragement and guidance during all of the ups and downs of my PhD journey. You believed in me when I was struggling and challenged me to help bring out the best in me. I am so grateful for all the time you have given in the past few years and for supporting me with all my decisions. Although not officially my supervisor I also want to thank Professor Martin Bunnage for all the supervision and input towards my PhD. A lot of the research would not have been possible without you, and you have inspired me to continue bridging the gap between research and clinical practice.

I have been fortunate enough to collaborate with a huge number of people during the PhD, all of whom have been incredibly supportive. Anna Brooke and Heather Tinkler helped me navigate NHS research and ethics procedures, which was no easy task but worth every moment. The clinical research would not have been possible without the help from all the staff at the NHS trust. Unfortunately, I cannot name them all, but I would like to give particular thanks the staff at the head injury rehabilitation unit: Angelita Cruz, Steph Davies, Jen Attwood, Aditi Sharma, Caroline Jones, Nicky Taylor, Abi, Jones, Jo Mason, Hannah Manning, Sarah Gibbin, Beth Watters, Beth Martin, Karen Smith, Carley Mead, Sarah Davies, Ellie McFarlane, Lian Foroudi, Saba Meky, Kirsty Nash, Kesta Purt, Nicky, Birdsey, Sarah Muddle, and Katie Thomas. They made me feel like part of the team and I would also like to acknowledge the incredible work they do with their patients. I am also grateful to all of the patients and their family members who took part in my study. Similarly, I am extremely grateful to all the families who took part in the ALSPAC study, the midwives for their help in recruiting them, and the whole ALSPAC team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists, and nurses.

My PhD was funded by the Economic and Social Research Council through the South West Doctoral Training Partnership and I would like to thank them all of the incredible opportunities they have given me and for taking the time to listen and give advice whenever needed. Thanks also go to Professor Skye McDonald from the University of New South Wales and Professor Roy Kessels from Radboud University for the materials they made available to me and time they gave to support my research. From within my department thanks go to Dr. Eleanor Kennedy and Dr. Miriam Cohen for the support and

resources they provided to support my research. I am also grateful to everyone in the Tobacco and Alcohol Research Group and all the staff (past and present) in the School of Psychological Science. I have enjoyed working with so many of you over the years.

Jaz, Maddy, Steph, Kirsten, Elis, Andy, Will, Laura, Kayleigh, Nick and all of the Upstairs Bears and Priory Road Punks that have been part my PhD journey, you are all amazing! I am so happy that our PhD's brought us together, even though most of us are glad they are over. Bristol has been my home for many years now and I cannot begin to acknowledge everyone who has taken in interest in how I was doing during my PhD and encouraged me to continue. I appreciate every conversation and message sent and could not have gotten here without the support from all my incredible friends. I am especially grateful to Laura, who has been there every step of the way and kept me sane whilst writing this PhD. Finally, I want to thank my family for always being there, especially my parents. You have shaped who I am and the values you have taught me are reflected in this piece of work. Danke Ulrike und Helmut!

COVID-19 Statement

Several studies in this thesis were impacted by Covid-19 restrictions. The most significant disruption was for the clinical research study presented in Chapter 4 of this thesis. In March 2020 only about half of the data collection for that study had been completed. As participants for the study were being recruited within a healthcare setting all research activity was paused for approximately nine months. Participants who had already consented to taking part in the study were no longer able to participate as most of them were discharged by the time research activities resumed. The loss in recruitment time meant that sample size targets had to be reduced to allow for completion of the study within the timeframe of this PhD. Once research activities were allowed to resume the Covid-19 restrictions meant that it was not feasible to conduct face to face testing sessions with patients. This meant that tasks and questionnaires were recoded for online presentation and study procedures had to be adapted and approved by the relevant ethics committee. By necessity control participants for this study were recruited online as opportunities for face-to-face testing were limited even outside of a healthcare setting. These changes in procedure halfway through data collection and switch from in person to online testing may have had an impact on data quality, although data checks suggest that in person and online performance are comparable in the TBI group. Due to the prolonged period of uncertainty around whether the clinical research study would be able to continue and be completed within the time frame of the PhD I decided to replicate the study design for the clinical study outside a healthcare setting. The online study presented in Chapter 5 is the result of this replication.

Another impact of the Covid-19 pandemic was that I was unable to complete a three month long institutional visit to the University of New South Wales in Sydney. I was due to leave for the visit in March 2020 to spend some time working in the lab run by Professor Skye McDonald. She is an expert in the field of social cognition in a TBI population and helped developed The Awareness of Social Inference Test (TASIT). The plan was to collaborate on the construct validation study presented in Chapter 3 of this thesis, which includes one of the subtests from TASIT. The plan had been to conduct the study online, so that aspect of the study was not impacted by Covid-19, but I was not able to benefit from the experience within that lab to inform my study design.

Author’s Declaration

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's *Regulations and Code of Practice for Research Degree Programmes* and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ALSPAC Avon Longitudinal Study of Parents and Children
BISI Brain Injury Screening Index
BPAQ Buss-Perry Aggression Questionnaire
CANTAB Cambridge Automated Neuropsychological Test Battery
CIS-R Clinical Interview Schedule – Revised
CL Common Language Effect Size
CT Computerised Tomography
DASS Depression Anxiety and Stress Scale
DSB Digit Span Backwards
EBT Emotion Bias Task
ERT Emotion Recognition Task
EET Emotion Evaluation Test from The Awareness of Social Inference Test
FEEST Facial expressions of emotion: Stimuli and Test
fMRI functional Magnetic Resonance Imaging
GAD Generalised Anxiety Disorder
GCS Glasgow Coma Scale
GFMT Glasgow Face Matching Test
ICC Intraclass Correlation Coefficient
MRI Magnetic Resonance Imaging
PTA Post Traumatic Amnesia
PTSD Post Traumatic Stress Disorder
RCT Randomised Control Trial
SEA The Social Cognition and Emotional Assessment
SST Stop Signal Task
SSRT Stop Signal Reaction Time
TAS Toronto Alexithymia Scale
TASIT The Awareness of Social Inference Test
TASIT-S Short version of The Awareness of Social Inference Test
TBI Traumatic Brain Injury

Chapter 1 Introduction

1.1 Thesis Motivation

My PhD project has investigated emotion recognition after Traumatic Brain Injury (TBI) using the Bristol Emotion Recognition Task (ERT). There is substantial evidence that emotion recognition is impaired after moderate to severe TBI (Babbage et al., 2011; Murphy et al., 2021), although less is known about emotion recognition after mild TBI (Theadom et al., 2019). I first became interested in this field during my Masters in Applied Neuropsychology, when I learnt that TBI can often lead to breakdown in relationships and social isolation (Ponsford et al., 2014). I hope that by better understanding changes in social cognition associated with TBI, such as the decrease in emotion recognition, we can begin to mitigate some of these difficulties.

It is striking that despite growing awareness of deficits in emotion recognition after TBI, social cognition is not routinely assessed in a clinical setting (Kelly et al., 2017). I think that bridging this gap between research and clinical practice is important, which means that we need to start developing emotion recognition tasks used in research for use in neuropsychological assessment (Kessels, 2019). Therefore, I chose to evaluate the utility of the Bristol ERT as a measure of emotion recognition in a TBI population. The task is already available as part of the Cambridge Automated Neuropsychological Test Battery but prior to this thesis had not been evaluated in this population. To effectively interpret emotion recognition performance on the Bristol ERT in a clinical setting it is also important to understand the impact of potentially confounding factors. Findings from my masters research indicated that clinical levels of anxiety could be associated negative bias on the Bristol ERT (Müller-Glodde, 2015). Concurrent research investigating the impact of state anxiety on emotion recognition measured using the Bristol ERT suggested that anxiety can also lead to deficits in emotion recognition (Attwood et al., 2017). This led me to wonder about associations between emotion recognition, TBI, and anxiety, which is why I decided to focus on anxiety as a potential confounder for emotion recognition in a TBI population in this thesis. Ultimately, I hope that the Bristol ERT will be a useful tool to assess emotion recognition and can be used to inform interventions for people struggling with social interactions after TBI.

1.2 Traumatic Brain Injury

Traumatic Brain Injury (TBI) is the modern terminology for a head injury and has been defined as “an alteration of brain function, or other evidence of brain pathology, caused by an external force” (Menon et al., 2010). Alteration in brain function can mean several things, including loss of consciousness at the time of the accident, post traumatic amnesia, and changes in mental or physical state due to neurological changes.

1.2.1 Prevalence

TBI is often referred to as the silent or hidden epidemic and is the leading cause of disability from neurological disease (Dewan et al., 2019; Maas et al., 2017; Majdan et al., 2016). In 2018, the headway brain association reported almost 156,000 hospitalisations for head injury in the UK alone (Headway, 2018). It is difficult to establish incidence of TBI, as many people with mild TBI do not seek medical attention and data is not always reliably available. Global incidence has been estimated at over 50 million new cases each year (Maas et al., 2017), with Dewan and colleagues (2019) modelling a worldwide incidence of 64 to 74 million cases annually. A meta-analysis conducted in 2013 suggested the prevalence of TBI in the general population to be around 12% (Frost et al., 2013), although Maas and colleagues (2017) have stipulated that due to increasing incidence and underreporting it is possible that 50% of people experience a TBI in their lifetime. Men are 1.5 to 2 times more likely than women to have a TBI and incidence is higher in young individuals, although high income countries are reporting increasing numbers of TBIs in older populations (Headway, 2018; Maas et al., 2017; Maas et al., 2008; Peeters et al., 2017).

1.2.2 Diagnosis

A range of indicators are used to give a diagnosis of TBI, including length of loss of consciousness, presence of Post Traumatic Amnesia (PTA), scores on the Glasgow Coma Scale (GCS), and other reported symptoms (McKee & Daneshvar, 2015; Menon et al., 2010). The GCS was developed by Teasdale and Jennett (1974, 1976) as a standardised measure of responsiveness after suspected TBI, and scores are based on eye, motor, and verbal reactivity to stimuli. Neuroimaging techniques like Computerised Tomography (CT) and Magnetic Resonance Imaging (MRI) can be used to inform a diagnosis of TBI but it remains a clinical decision (Schweitzer et al., 2019).

The Mayo Classification for TBI Severity (Mayo system) was developed to help standardise clinical decision making about severity of TBI based on all available indicators (Malec et al., 2007). There are three categories which capture the likelihood that a TBI has occurred: 1) moderate to severe TBI (definite TBI), 2) mild TBI (probable TBI), and 3) possible TBI. Indicators used are the length of loss of consciousness, PTA, scores on the GCS, neuroimaging abnormalities, and neurological symptoms (Friedland, 2013; Malec et al., 2007; Teasdale & Jennett, 1974, 1976). Whilst this sounds relatively straight forward it can be difficult to accurately assess severity because reports of loss of consciousness, or PTA are difficult to estimate if there are no witnesses to the event (Ruff et al., 2009). Further, intoxication and shock might result in similar symptoms and lead to lower GCS scores. Secondary brain injury, (for example, lack of oxygen to the brain due to hypoxia or hypotension, or intracranial pressure due to swelling or build-up of fluid in the brain) can also have a big impact on outcomes after TBI (Friedland, 2013; Maas et al., 2008). Additionally, there is substantial overlap in self-reported symptoms that occur post TBI with symptoms reported after injuries that do not involve trauma to the head (Gasquoine, 2020). These complexities have led to inconsistencies in diagnosis and classification of TBI. Many overlapping definitions exist that are revised as understanding develops.

1.2.2.1 Moderate to Severe Traumatic Brain Injury

A diagnosis of moderate to severe TBI is given when there is no doubt that there has been damage to the brain. The Mayo system classification for moderate to severe TBI is defined as having one or more of the following criteria: GCS score of less than 13 (Moderate 9-12, Severe 3-8) within 24 hours of the TBI event, loss of consciousness of more than half an hour, and PTA for 24 hours or longer (Azouvi et al., 2017; Friedland, 2013; Malec et al., 2007). Additionally, any external or neuroimaging evidence of significant damage to the brain would likely lead to a diagnosis of moderate to severe TBI. It is important to note that whilst everyone who matches these criteria has definitely experienced a TBI, this does not mean that it is a homogeneous group, and outcomes can vary drastically (Friedland, 2013; Svingos et al., 2019).

1.2.2.2 Mild Traumatic Brain Injury

The aim of any definition of mild TBI is to exclude any injuries that are moderate or severe, as well as excluding any head injuries that did not result in an injury to the

brain (Gasquoine, 2020). The most commonly used criteria for mild TBI has been loss of consciousness for less than 30 minutes (Carroll et al., 2004; Ruff et al., 2009). However, Gasquoine (2020) argues that presence of PTA (less than 24 hours long) and not loss of consciousness should be used as an indicator to establish whether a mild TBI has occurred. The Mayo system classification suggests that a diagnosis of mild TBI should be given if one of the following criteria is met: any loss of consciousness up to 30 minutes, presence of PTA that lasts no longer than a day, or a skull fracture that leaves the brain dura intact (Friedland, 2013; Malec et al., 2007). In terms of the Glasgow Coma Scale, mild TBI is defined as a score of 13 to 15 (Azouvi et al., 2017; Teasdale & Jennett, 1976). It is often difficult to establish whether a head injury has resulted in damage to the brain, so if indicators of mild TBI are not available or unclear but there are changes in brain function a diagnosis of possible TBI can be given. Possible TBI is defined as presence of neurological symptoms such as blurred vision, confusion, dizziness, headache, nausea, being dazed or other focal neurological symptoms (Friedland, 2013; Malec et al., 2007). These neurological symptoms are also included as diagnostic criteria for mild TBI and are commonly associated with having a concussion (Levin & Diaz-Arrastia, 2015). In fact, concussion and mild TBI are often used interchangeably in the literature (Gasquoine, 2020; Levin & Diaz-Arrastia, 2015; Ruff et al., 2009). It is important to recognise that these categories are representative of the level of confidence that a TBI has occurred and should not be treated as distinct conditions.

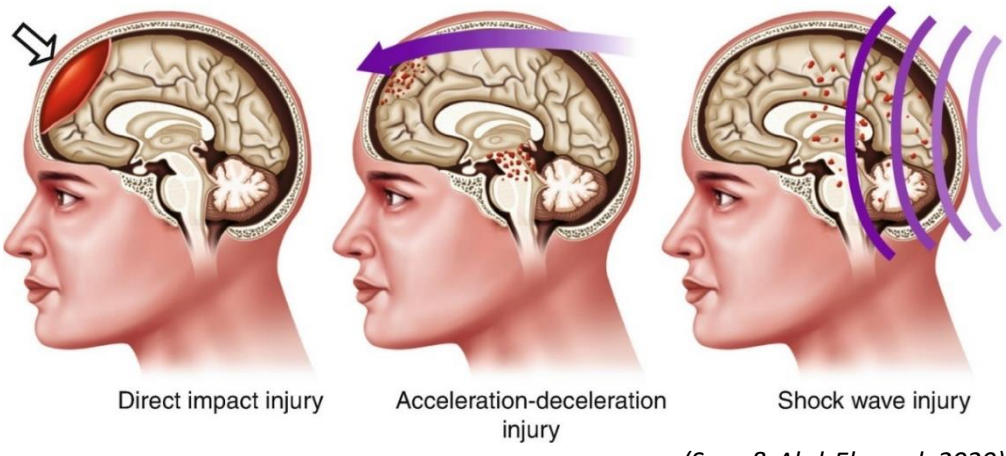
It is difficult to accurately diagnose mild TBI because of the inability to accurately identify structural damage, and the comorbidity of symptoms with other conditions. As a result, there is large variation in the outcomes for people diagnosed with mild TBI (Gasquoine, 2020; Ruff et al., 2009). CT scans do not reliably predict severity of symptoms or outcomes after TBI, which is one of the reasons they are not used as the primary diagnostic tools to define severity of TBI (Brown et al., 2019; Fure et al., 2021). Recent developments in brain imaging have shown that some types of brain damage are not detectable using traditional CT or MRI scanning (Bigler, 2013). Development of non-invasive diagnostic methods for TBI such as Diffusion Tensor Imaging could support clarification of classification in years to come but are not yet available for routine clinical assessments (Bigler, 2013; Joseph et al., 2018). Si and colleagues (2018) used a data driven approach to identify sub-classes of mild TBI with distinct outcome profiles. This shows that there are likely to be different categories of mild TBI that we are currently unable to identify effectively.

1.2.3 Pathology

TBI injuries often involve both primary and secondary trauma (Azouvi et al., 2017; Ng & Lee, 2019). Primary trauma is the damage caused by the initial impact of force and secondary traumas are systemic complications that arise because of the impact. Examples of secondary trauma are vascular injury, lack of oxygen due to hypoxia or hypotension, swelling, or inflammation (McKee & Daneshvar, 2015). These have the potential to cause further damage but can sometimes be stopped or reversed given appropriate treatment. The primary trauma is not reversible and the type of primary impact, as well as the force, direction, and duration of the impact influence the type and extent of damage caused (Maas et al., 2008; McAllister, 2011). An injury is classified as open if an external object has penetrated the brain and closed if an external force has caused damage without penetrating the brain. There are three types of external forces that can cause damage to the brain; direct impact, a rapid acceleration/deceleration, or a blast (Figure 1-1; Azouvi et al., 2017; Maas et al., 2008; McAllister, 2011; Ng & Lee, 2019; Suer & Abd-Elsayed, 2020). Most people experience a combination of these different types of primary impact, adding to the complexity of diagnosis and treatment (McAllister, 2011).

Figure 1-1

Different forces that can cause a TBI



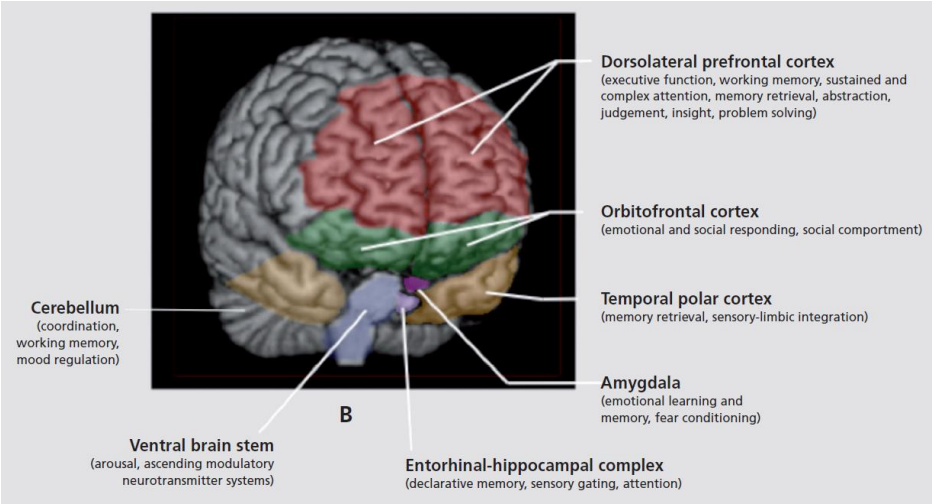
(Suer & Abd-Elsayed, 2020)

The damage caused by the primary impact can be focal (localised to one area) or more widespread, referred to as diffuse (Azouvi et al., 2017). Localised contusions are usually seen at the site of direct impact and areas where the brain has impacted with the skull

due to the force of the TBI (Ng & Lee, 2019). Given the mechanism of this type of injury there are certain regions of the brain more vulnerable to damage (Figure 1-2).

Figure 1-2

Brain areas that are vulnerable to damage after TBI



(McAllister, 2011)

The frontal and temporal regions of the brain are particularly at risk because the bone structure and ridges in the skull around those areas are more likely to cause damage to the brain when impacted (Bigler, 2007; McAllister, 2011). Rapid acceleration/deceleration forces are also likely to cause the brain to impact in those areas and can cause tearing of blood vessels that may result in focal or diffuse damage (McKee & Daneshvar, 2015). Acceleration/deceleration and rotational forces are very likely to cause diffuse axonal injury (shearing of neural axons), which is usually multifocal or diffuse, damaging white matter tracts in the brain (McKee & Daneshvar, 2015; Ng & Lee, 2019). A blast or shockwave can also cause axonal damage and is generally associated with diffuse damage to the brain (Ng & Lee, 2019). Diffuse axonal injury is present in the majority of TBI cases and is thought to be the main type of damage in mild TBI, whilst moderate to severe TBI is usually a combination of focal and diffuse damage (Calvillo & Irimia, 2020; McKee & Daneshvar, 2015; Ng & Lee, 2019). The degree of axonal shearing is thought to be linked to severity of TBI, but it requires MRI or Diffusion Tensor Imaging to detect, as it is not visible on CT scans (Azouvi et al., 2017; Ng & Lee, 2019; Wallace et al., 2018). Finally, TBI is also associated with changes in brain function detected using functional MRI, such as changes in the default mode network and altered connectivity in the frontal temporal networks (Azouvi et al., 2017; Yeates et al., 2017).

In summary, injury to the frontal-temporal regions of the brain as well as diffuse axonal damage are characteristic of TBI, but injury is not limited to those areas. Given the variable nature of injuries and possible secondary complications, outcomes after TBI are heterogenous and often difficult to predict (Calvillo & Irimia, 2020; Svingos et al., 2019). Improving understanding of psychosocial outcomes after TBI is important to develop appropriate interventions and support for people living with TBI.

1.2.4 Outcomes

Most people who experience a TBI survive and make a full recovery, but a substantial number live with long-term effects. A recent European based cohort study investigating outcomes six months post injury found that almost 90% of patients that were seen for a TBI survived (Steyerberg et al., 2019). Most people had fully recovered after six months, however around 25% of people seen only in ER and over 50% of people admitted to hospital reported ongoing difficulties (Steyerberg et al., 2019). A longitudinal study of 141 participants with mild to severe TBI found that over half the participants reported cognitive and behavioural changes even 10 years post injury (Ponsford et al., 2014). Having a moderate to severe TBI is more likely to be associated with long-term difficulties but an estimated 10% to 30% of cases categorised as mild also report persistent symptoms and disability (Maas et al., 2017; Prince & Bruhns, 2017; Steyerberg et al., 2019). This means that there are a growing number of people living with long-term outcomes of TBI and it is increasingly understood as a chronic health condition (Azouvi et al., 2017; Forslund et al., 2019; Galgano et al., 2017; Jourdan et al., 2018; Maas et al., 2017). Understanding the long-term outcomes associated with TBI is important to improve the treatment and support available.

Difficulties experienced after TBI include somatic symptoms, psychiatric consequences, behavioural changes, and cognitive deficits (Azouvi et al., 2017; Calvillo & Irimia, 2020; Prince & Bruhns, 2017). Commonly reported somatic symptoms are headaches, dizziness, fatigue, and sleep disturbances (Dikmen et al., 2010). There is an increased risk of anxiety and depressive disorders after injury (Alway et al., 2016; Ashman et al., 2004; Ponsford et al., 2018). Concerning behavioural changes include increased irritability or even aggression (Dikmen et al., 2010; Phyland et al., 2021), apathy (Worthington & Wood, 2018), and being socially inappropriate (Hicks et al., 2017). Cognitive deficits commonly associated with TBI are impaired memory, decreased processing speed and attention, and difficulties with executive functioning (Azouvi et al.,

2017; Draper & Ponsford, 2008). More recently, changes in social cognition have also been associated with TBI and are emerging as an important factor in brain injury outcome and rehabilitation (Allain et al., 2019; Calvillo & Irimia, 2020).

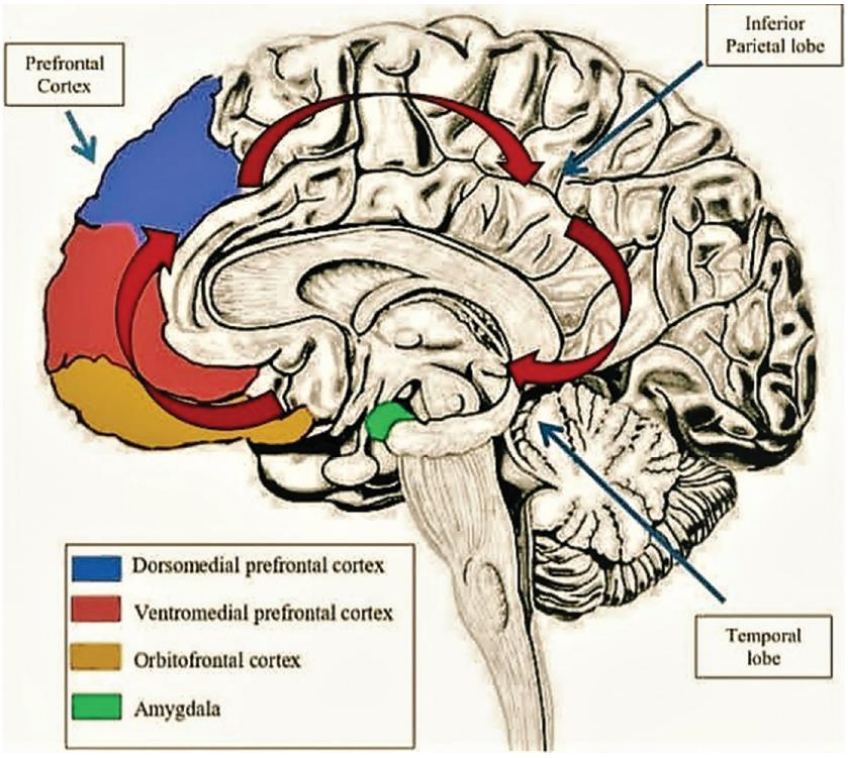
1.2.4.1 Social cognition

Social cognition refers to the processes underlying our ability to perceive, interpret, and understand social information that is required for appropriate social interaction (Allain et al., 2019; Cassel et al., 2019). Cognitive aspects include theory of mind, perspective taking and cognitive empathy; whilst affective aspects include emotion perception and affective empathy. Additionally, social and emotional self-awareness plus self-regulation are considered important aspects of social cognition that could underpin some of the other cognitive and affective aspects (Cassel et al., 2019).

Having a TBI is associated with difficulties across many if not all aspects of social cognition (Allain et al., 2019; Cassel et al., 2019; Maggio et al., 2020; McDonald, 2013) and there is overlap between the brain areas associated with social cognition (Figure 1-3) and the brain areas vulnerable to damage due to TBI (Figure 1-2).

Figure 1-3

Brain areas associated with social cognition



(Maggio et al., 2020)

Moderate to severe TBI has consistently been linked to deficits in emotion recognition (Babbage et al., 2011) and theory of mind (Lin et al., 2021). There is also evidence for decreased empathy and increased alexithymia i.e. difficulties in identifying and thinking about emotions (de Sousa et al., 2010; Williams & Wood, 2010). To date, little is known about social cognition after mild TBI (Calvillo & Irimia, 2020; Theadom et al., 2019). In a longitudinal follow-up study Theadom and colleagues (2019) found evidence for social cognition deficits in about 20% of participants with mild TBI four years post injury. There is also evidence for changes in brain activation when emotional stimuli are presented to people with persisting symptoms after mild TBI (Bohorquez-Montoya et al., 2020). Further studies are needed to develop our understanding of changes in social cognition after mild TBI.

There is an increased understanding of changes in social cognition after TBI and the impact it can have on patient outcomes. TBI can have a long-term impact on social relationships and communication (Ponsford et al., 2014) and there is evidence that social cognition might be underlying successful social functioning (May et al., 2017; Milders, 2019; Milders et al., 2008). Unlike cognitive functions such as memory and attention, social cognition is not consistently assessed in a clinical setting (Kelly et al., 2017; Maggio et al., 2020). This is problematic because arguably there is only a weak link between general cognitive deficits and social outcomes (Milders, 2019). It is important to better evaluate social cognition after TBI, to help inform interventions and improve quality of life for people living with TBI. This thesis aims to further our understanding of social cognition after TBI by using the Bristol Emotion Recognition Task (described in section 1.3.2) to investigate emotion recognition in this population.

1.3 Emotion Recognition

Emotion recognition is the process by which an emotion expressed by another person is accurately perceived and identified (Adolphs, 2002b; Beer & Ochsner, 2006; Knapp et al., 2013). Non-verbal cues such as body language, vocal inflections, and facial expressions are an important part of social communication because they provide information about other people's emotional states or social intentions (Aviezer et al., 2008; Dukes et al., 2017; Knapp et al., 2013; Neumann, Keiski, et al., 2014; Parkinson, 2005). Facial expressions, in particular, are a rich source of information about emotions and are used as a means of social communication (Knapp et al., 2013; Parkinson, 2005).

In the mid to late twentieth century, Basic Emotion Theory became a foundation for the study of emotion recognition from facial expressions (Crivelli & Fridlund, 2019; Ekman, 1992; Ekman et al., 1987; Keltner, Sauter, et al., 2019). Ekman and colleagues introduced what are known as the six 'basic' emotions, namely happiness, sadness, anger, fear, surprise, and disgust (Crivelli & Fridlund, 2019; Ekman, 1992; Friesen & Ekman, 1976). A common criticism is that this categorisation is overly simplistic and to date there is no consensus on the definition of an emotion (Crivelli & Fridlund, 2019; Keltner, Tracy, et al., 2019). Some of the ways that emotions are conceptualised are as behavioural states (i.e. withdrawal or approach), moods (e.g. anxious or content), social emotions (e.g. pride or guilt), and of course 'basic' emotions (Adolphs, 2002b; Crivelli & Fridlund, 2019). There is evidence for distinct cross cultural facial expressions that map onto four of the 'basic' emotions, happy, sad, angry/disgusted, and fearful/surprised (Jack et al., 2014; Jack et al., 2016), and it is possible that there are distinct neural correlates for 'basic' emotions (Celeghin et al., 2017). So, despite 'basic' emotions not capturing the full range of emotions used in social interaction, an inability to identify those emotions could be indicative of wider difficulties.

Linked to the question of how an emotion is defined is the question of how it is accurately recognised. There is ongoing debate about what processes are involved in emotion recognition and whether processes are more bottom-up, i.e., perceptual, or top-down, i.e., inferential (Adolphs, 2002b; Joseph & Newman, 2010; Newen et al., 2015). It is possible that emotion recognition is a purely perceptual process, requiring categorisation a facial properties and retrieval of an appropriate label, with little to no understanding of emotions required (Adolphs, 2002b; Newen et al., 2015). Equally, emotion recognition could be inextricably linked to the ability to identify and express one's own emotions and dependent on emotional understanding (Adolphs, 2002b). There is no one coherent model of emotion recognition and its function in the wider context of social cognition or emotional intelligence. Probably the most comprehensive theories are Basic Emotion Theory and related models, which have continued to develop to reflect current understanding (Crivelli & Fridlund, 2019; Keltner, Tracy, et al., 2019). An example of a different model is the framework of Emotion Understanding in Recognition and Knowledge Abilities (EUReKA) introduced by Castro and colleagues (2016) . They propose that emotion recognition and emotion knowledge together constitute emotional understanding, and argue that there are four aspects of emotion recognition, of which labelling prototypical (i.e., basic) expressions is one.

In sum, although there is no consensus on processes underlying emotion recognition, the ability to accurately identify 'basic' emotions is recognised as one component of emotion recognition that potentially acts as a foundation for more complex processes (Adolphs, 2002b; Castro et al., 2016; Celeghin et al., 2017; Jack et al., 2014). It is important to acknowledge the complexity involved in understanding the process of emotion recognition, however this thesis does not aim to further develop these theories. Instead, it focuses on what an emotion recognition task using the six 'basic' emotions can tell us about difficulties in emotion recognition after TBI. The ability to recognise emotions from facial expressions has been linked to successful social interaction (Avery et al., 2016; Milders et al., 2008). Given this association between emotion recognition and social interaction, these admittedly simplified/basic emotion recognition tasks could provide useful information for diagnosis and treatment of social cognition deficits after TBI.

1.3.1 Measuring Emotion Recognition

Despite an increased understanding of changes in social cognition after TBI and the impact it can have on patient outcomes, it is not consistently assessed in clinical practice (Kelly et al., 2017; Maggio et al., 2020). One reason is that tests of social cognition are not yet well established, and many tasks used for research lack validation and information about reliability (Howieson, 2019; Kelly et al., 2017). D'Souza and colleagues (2019) argue that even tasks that are well established in clinical practice do not always have sufficient information about validity and reliability in a TBI population. A better understanding of psychometric concepts, such as reliability, validity, and discriminability (Cooper et al., 2017) is crucial for both research and clinical practice. Simply put, reliability refers to consistency in measurement given stable conditions (Bowden, 2017; Streiner et al., 2015). Reliability coefficients can be used to inform study design and analysis, and should be consistently reported for a more effective interpretation of experimental results (Parsons et al., 2019). In clinical practice this translates to providing context for a clinical assessment that goes beyond simple test scores (Kessels, 2019; Meyer et al., 2001). A recent study by Camargo and colleagues (2018) defined validity as "the degree to which collected evidence, theory and argumentation support inferences based on observed scores". In other words, an assessment of whether valid inferences can be made based on the outcome measures of a particular task or questionnaire. In both research and clinical practice, it is important to check that the inferences being made have a good foundation, otherwise conclusions

drawn are not useful and potentially misleading. Discriminability refers to the ability to distinguish between two groups based on the observed scores on a task (Streiner et al., 2015). This is particularly important in a clinical setting if a task is being used to inform diagnosis but is also used to investigate the effectiveness of interventions, and group differences in an experimental setting.

In the area of Neuropsychology, Howieson (2019) identified three test batteries of social cognition that have been validated and provide normative data; The Social Cognition and Emotional Assessment (Funkiewiez et al., 2012), The Awareness of Social Inference Test (TASIT, McDonald et al., 2003), and the EMOTICOM battery (Bland et al., 2016). The Social Cognition and Emotional Assessment was developed specifically for detection of frontotemporal degeneration and has not been utilised in a TBI setting. TASIT was specifically developed to assess social cognition after TBI and has been recommended as the most appropriate tool to assess social cognition in a moderate to severe TBI population (Honan et al., 2019). The EMOTICOM battery was not designed with a particular population in mind and to date has not been evaluated or used in a TBI setting. The above batteries each include a different emotion recognition task. All three tasks are so called forced choice labelling tasks using the six basic emotions, which means that participants are presented with a stimulus and are asked to identify the emotion presented using the labels provided. Where the tasks differ is in the type of stimuli used.

1.3.1.1 Emotion Recognition Tasks

Paiva-Silva and colleagues (2016) published a systematic review of tasks used to assess emotion recognition from facial expressions over the past 20 years. The review showed that the most common methodology used to investigate behavioural emotion recognition is forced choice labelling tasks, followed by emotion matching tasks. They reported that the stimuli themselves vary dramatically across studies, and could be categorised as static or dynamic, natural versus manipulated, and photographic, computer generated, or drawn. Natural static stimuli were identified as the most used form of stimuli, despite criticisms that facial expressions in real life are often more ambiguous and present dynamically (Krumhuber et al., 2013). Morphed sequences both in static form or as dynamic clips have been used to address this issue and help avoid ceiling effects in performance (Paiva-Silva et al., 2016; Suzuki et al., 2006). Arguably, dynamic morph sequences have been shown to have a slight advantage in accuracy and

speed of recognition compared to static stimuli (Calvo et al., 2016). The emotions most commonly investigated are the 'basic' emotions of happiness, sadness, fear, anger, disgust and surprise (Ekman & Cordaro, 2011). The number of tasks available to evaluate emotion recognition can make it difficult to draw comparisons between studies (Paiva-Silva et al., 2016).

The first and probably most well-known set of emotion stimuli was developed by Ekman and Friesen (Ekman & Friesen, 1978; Friesen & Ekman, 1976). The Pictures of Facial Affect are photographs of facial expressions for happiness, sadness, anger, fear surprise, and disgust, as well as a neutral expression. They are still commonly used despite other data sets such as the Karolinska Directed Emotional Faces having been developed that are arguably more comprehensive (Goeleven et al., 2008; Young et al., 2002). It is not surprising that in a scoping review of social cognition measures used after acquired brain injury, Wallis and colleagues (2021) identified the Pictures of Facial Affect as the most widely used measure of emotion recognition. Although the same basic stimuli are used, studies vary in how the stimuli are prepared and presented. Examples of validated forced choice labelling tasks using the Pictures of Facial Affect are the Ekman 60 Faces test from The Facial Expressions of Emotion: Stimuli and Test (Young et al., 2002) and the Identification of emotions test used in the Social Cognition and Emotional Assessment battery (Funkiewiez et al., 2012).

The other commonly used measure of emotion recognition after acquired brain injury identified by Wallis and colleagues (2021) was the Emotion Evaluation Test (EET) from TASIT (McDonald et al., 2003) and its short version TASIT-S (Honan et al., 2016). Unlike the tasks using the Pictures of Facial Affect, the EET from TASIT does not use facial expression stimuli. Instead, participants are presented a series of video clips showing one or two actors in dialogue. The scripts used are neutral, but participants are asked to identify the emotion presented based on various nonverbal cues, including facial expressions. After viewing the vignette participants are asked to identify the emotion presented as one of the six 'basic' emotions or as neutral. TASIT is currently one of only two outcome measures of social cognition recommended for use after moderate to severe TBI (Honan et al., 2019), and the only measure that includes an emotion recognition task. This highlights the need for further development of tasks in this domain and the emotion recognition task part of the EMOTICOM battery could prove useful for neuropsychological assessment (Howieson, 2019).

The EMOTICOM battery was developed as a comprehensive neuropsychological test battery of affective (i.e., social) cognition (Bland et al., 2016). The emotion recognition task included in this battery was a four emotion version of the Emotion Recognition Task from the Cambridge Automated Neuropsychological Test Battery (CANTAB) owned by Cambridge Cognition Ltd. The task was developed by Professors Marcus Munafò and Ian Penton-Voak at the University of Bristol and will be referred to in this thesis as the Bristol Emotion Recognition Task (ERT). It is described in detail in the next section. The EMOTICOM battery has not been evaluated in TBI setting so the utility of the Bristol ERT in this population has not yet been established.

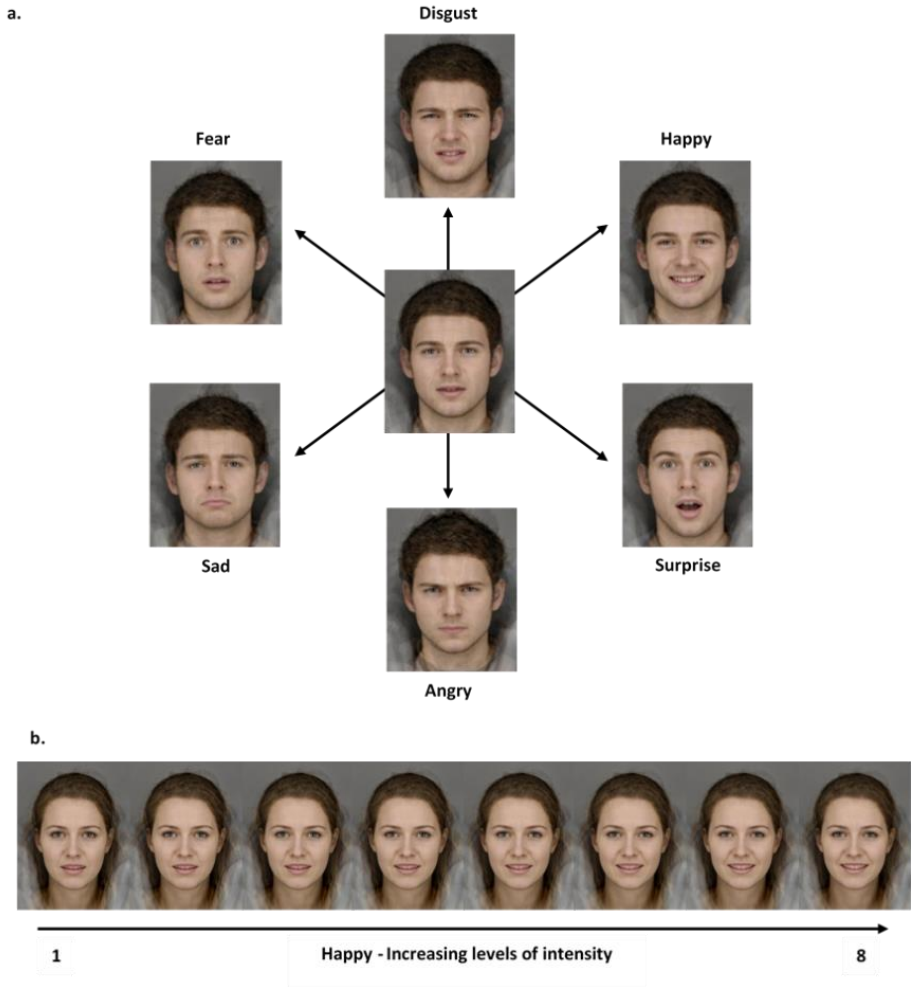
1.3.2 Bristol Emotion Recognition Task

The Bristol ERT is a forced choice labelling task using 354x464 pixel front view photographs of facial expressions as the stimuli. The emotions available are the six 'basic' emotions, happy, sad, anger, fear, surprise, and disgust. Instead of selecting a set of photographs capturing a facial expression, the stimuli were created by merging photographs of 12 people making the facial expression for that emotion. Figure 1-4 shows the male and female morphed stimuli developed for this task. Computer generated morphed stimuli have several benefits over photographs such as the Pictures of Facial Affect, including the reduction of idiosyncrasies in facial expressions and the ability to manipulate expression intensity (Vetter & Walker, 2011). Judgements made about a face have been shown to influence the ability to identify the emotion expressed (Colonnelle et al., 2019; Vuilleumier & Righart, 2011), and using a standardised computer generated face could help reduce variability in responses based on judgments about the stimuli presented. A commonly used method to generate emotional expressions at different levels of intensity is to create morph sequences consisting of a series of images that are a combination of two endpoint images. The aim is to increase sensitivity of emotion recognition tasks and prevent ceiling effects (Paiva-Silva et al., 2016). For the Bristol ERT, a linear morph sequence was used to generate stimuli with different levels of intensity for each emotion. The lowest intensity level is highly ambiguous, showing very little of the target emotion and thus only marginally different from the baseline image. The highest intensity level is completely unambiguous, showing the target emotion at 100% intensity. The remaining levels of intensity consist of equal increments between those two levels (See Figure 1-4 *b* for an example of an 8-face sequence). Single emotion morph sequences are often created by merging a neutral expression with an emotion expression. However, using a neutral face as a baseline is

potentially problematic because neutral faces can be perceived as threatening (Yoon & Zinbarg, 2008). Arguably responses for low intensity stimuli could skew towards being perceived as angry. Consequently, the stimuli for the Bristol ERT were generated using a prototypical non emotional face as a baseline instead of a neutral facial expression. The prototypical face is a composite of the 100% intensity stimuli modified to contain equal proportions of six 'basic' emotions (See central picture in section *a* of Figure 1-4). Israelashvili and colleagues (2021) found that prototypical faces are an appropriate proxy for natural facial expressions and arguably they are a more appropriate as a baseline for morph sequences than neutral facial expressions (Skinner & Benton, 2010).

Figure 1-4

Example of stimuli used in the Bristol Emotion Recognition Task



a. Example of male stimuli depicting all six emotions at 100% intensity around the prototypical face, which consists of equal proportions all these six emotions.
b. Example of female stimuli in a sequence showing happiness at eight levels of intensity from lowest to highest

(Müller-Glodde, 2015)

A common criticism of emotion recognition tasks is that they are not realistic because facial expressions develop dynamically and asking participants to make a judgement by looking at a static image is not an adequate representation of the emotion recognition process (Calvo et al., 2016; Darke et al., 2019). Some tasks use dynamic morph sequences to address this issue (Montagne et al., 2007). An alternative adaptation to better reflect real life perception could be to present the stimuli for only a short period of time, because sometimes expressions of emotion can be very brief. Humans can pick up on these brief cues with astounding accuracy (Smith, 2012; Sweeny et al., 2013). Both experimental and neuroimaging evidence indicate that emotion recognition from facial expressions is thought to occur through a combination of automatic (bottom-up) processes and top-down modulation (Vuilleumier & Righart, 2011). Having a reduced presentation time for the stimuli could help tap into the more automated cognitive process. Consequently, the presentation time for stimuli on the Bristol ERT is usually 500ms, as information to accurately categorise emotions is thought to be available after 200ms (Schyns et al., 2009). In conjunction with the short presentation time on the Bristol ERT, a mask is included directly after presentation to avoid afterimage effects. The aim is to minimise the possibility of adaptation effects that could impact responses on subsequent trials, as Webster and colleagues (2004) found that adaptation to a particular expression can influence the judgements made about subsequent ambiguous expressions. In sum, the stimuli used and the way they are presented aim to reduce trial to trial noise, which should help decrease unwanted variation that could obscure potential emotion recognition effects.

The Bristol ERT can be used to measure emotion recognition accuracy as well as provide information about response biases that could be indicative of interpretive biases. As an overall measure of accuracy, 'total hits' (number of correct identifications) is calculated. This is the primary outcome measure used throughout this thesis. In addition, a 'hit rate' and 'false alarm rate' (number of times an emotion was correctly or erroneously chosen, respectively) is recorded for each emotion. The hits and false alarms for each emotion can be used to calculate an unbiased hit rate based on a formula presented by Wagner (1993).

$$H_u = \frac{hits^2}{16 \times (hits + false\ alarms)}$$

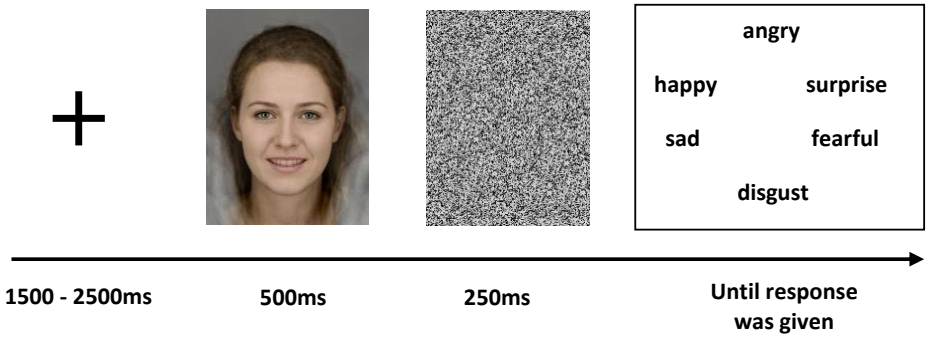
Higher proportion of total hits or H_u for each emotion indicate increased emotion recognition accuracy on the Bristol ERT. The total number of times an emotion was selected (number of hits plus false alarms for each emotion) can be used as a measure of response bias. A response score for a particular emotion could indicate an interpretive or perceptual bias towards that emotion. The Bristol ERT has been used successfully in different areas of research to investigate the association between emotion recognition and alcohol use (Eastwood et al., 2020; Freeman et al., 2018), anxiety (Attwood et al., 2017), autism (Griffiths et al., 2019), bipolar disorder (Russo et al., 2015), and depression (Bamford et al., 2015). It was also used to investigate emotion recognition after acquired brain injury as part of a master's project (Müller-Glodde, 2015), which was a feasibility study for use of the Bristol ERT in a TBI population.

1.3.2.1 Description of the task used throughout this thesis

In the version of the Bristol ERT used in this thesis, all six emotion stimuli were presented at eight levels of intensity. Both male and female stimuli were used, and all the faces presented were Caucasian. This means a total of 96 stimuli were included and presented once to each participant. The order of presentation was randomised, and participants were offered a break after completing 48 trials. Each trial started with the presentation of a fixation cross (1500 to 2500ms), followed by presentation of the face stimulus for 250 to 500ms and then a visual noise mask for around 250ms (See Figure 1-5 for an example of a trial). Participants were presented with a response screen showing the six emotions included in this task and asked to select the emotion that best described the facial expression they had seen. There is no time constraint on giving a response, but participants are instructed to respond as quickly and accurately as they can. Once a response was selected the next trial started automatically.

Figure 1-5

Trial sequence for Bristol Emotion Recognition Task



1.4 Emotion Recognition and Traumatic Brain Injury

There is consistent evidence of impaired emotion recognition after TBI across a range of emotion recognition tasks (e.g. Knox & Douglas, 2009; McDonald et al., 2011; Milders et al., 2003; Rosenberg et al., 2014; Rosenberg et al., 2019; Zupan et al., 2014). In 2011, Babbage and colleagues conducted a meta-analysis of studies investigating emotion recognition after moderate to severe TBI using static facial emotion recognition tasks. The results indicated that participants with TBI are worse at emotion recognition tasks compared to neurologically healthy controls and that the effect size for this deficit is large (Hedges' g = 1.1, 95% CI 0.97 to 1.25). An estimated 50% of people with TBI present with emotion recognition difficulties based on a cross-sectional study conducted by Biszak and Babbage (2014). Murphy and colleagues (2021) recently published an updated systematic review and meta-analysis of emotion recognition deficits after TBI, including forced choice labelling tasks for basic emotion across a range of presentation modalities. They calculated an unbiased effect size estimate of Hedges' g = 0.79 (95% CI 0.61 to 0.96), which is slightly smaller than the one reported by Babbage and colleagues but still a moderate to large effect of emotion recognition deficits after TBI. They did not find a difference in effect across task modalities and reported that the two most used tasks included were the Pictures of Facial Affect and the Emotion Evaluation Task from TASIT. Murphy and colleagues (2021) did not actively exclude studies looking at mild TBI, but were unable to find suitable studies to be included in their analysis, highlighting a clear gap in the literature.

There are few longitudinal studies of emotion recognition after TBI for an adult population, but evidence available indicates that deficits in emotion recognition after TBI are persistent. Ietswaart and colleagues (2008) recruited participants shortly after a TBI or orthopaedic injury and assessed emotion recognition around two months post injury and again one year later. TBI participants performed worse than orthopaedic injury controls at both time points and there was no evidence for improvement specific to the TBI group. Theadom and colleagues (2019) assessed long term impact of mild TBI on emotion recognition. They reported a trend towards decreased emotion recognition after mild TBI but there was no evidence for a difference in performance between participants with mild TBI and healthy controls four years post injury. Initially there was evidence that negative emotions were specifically impaired (Croker & McDonald, 2005) but more recent research indicates that TBI patients are impaired on all emotions (Murphy et al., 2021; Rosenberg et al., 2018). This change could reflect the development

of emotion recognitions tasks, with newer tasks being able to address issues such as the ceiling effects observed for recognition of happiness. This shows the importance of using well developed and validated tasks to effectively measure and interpret changes in emotion recognition. Murphy and colleagues (2021) did suggest that based on the magnitude of difficulties observed for each emotion, that negative emotions such as fear and anger are associated with a larger impairment than happiness and surprise.

A review of emotion recognition performance on behaviour outcomes after TBI suggested that better emotion recognition is associated with a better social/functional outcome (Milders, 2019). In turn, it is not surprising that deficits in emotion recognition have been associated with poorer outcome after TBI (Milders, 2019; Milders et al., 2008). For example, decreased emotion recognition and anger misattributions have been linked to other reported socially inadequate behaviour after moderate to severe TBI (Jorna et al., 2021; Spikman, Boelen, et al., 2013). Given that emotion recognition deficits after TBI are well established and there is evidence that these are linked to social functioning outcomes, it is difficult to understand why emotion recognition is not more routinely assessed and considered in clinical practice. Kelly and colleagues (2017) highlighted that there is a lack of awareness and clarity around the assessment of emotion recognition that needs to be addressed.

1.4.1 Brain structure and functioning

Emotion recognition is not associated with just one brain area but is comprised of a network of different brain areas and neural networks. Areas that have been linked to facial emotion recognition are the amygdala, thalamus, temporal /occipital brain structures (e.g., fusiform gyrus), as well as the medial and orbital frontal lobes (Adolphs, 2002a; Haxby & Gobbini, 2011; Henry et al., 2016; Rosen & Levenson, 2009). Some of these areas associated with emotion recognition are vulnerable to damage after TBI, including the orbitofrontal cortex and the amygdala (Bigler, 2007; McAllister, 2011). Spikman and colleagues (2012) found that poor emotion recognition after moderate to severe TBI is associated with damage to the orbitofrontal cortex and the amygdala has been associated with emotion recognition performance (Adolphs, 2010). Further, both focal and diffuse axonal injury, common after TBI, have also associated with changes in emotion recognition (Yassin et al., 2017). However, white matter damage is also associated with generally lower cognitive performance (Wallace et al., 2018) so the changes in emotion recognition could be the result of a more general cognitive deficit.

There is also evidence of changes in brain functioning related to emotion recognition after TBI. Xiao and colleagues (2017) conducted a meta-analysis of fMRI studies investigating social cognitive deficits after TBI. They found that deficits were linked to decreased activation in the tempo parietal junction and medial prefrontal cortex. Rigon and colleagues (2017) found that the ability to recognise emotions from facial expressions was associated with differences in resting state functional connectivity in participants with TBI compared to healthy controls. There is clearly evidence for a neural basis of emotion recognition deficits after TBI, but small sample sizes and lack of specificity in a lot of studies means that it is not possible to make any causal inferences based on current research.

1.4.2 Factors that could influence Emotion Recognition after Traumatic Brain Injury

When considering changes in emotion recognition after TBI it is important to acknowledge that there are a lot of factors that could be moderating or mediating these changes. General cognitive ability could influence the ability to complete the emotion recognition tasks, or inability to effectively perceive faces could be impacting emotion recognition performance (J. D. Henry et al., 2015). Arguably, emotion recognition is linked to understanding of emotions, so it is also possible that difficulties in identifying and describing emotions (alexithymia) could underlie performance on emotion recognition tasks (Williams & Wood, 2010). There is also evidence that age and sex impact emotion recognition (Abbruzzese et al., 2019). Men are more likely to have a head injury and certain age groups are at a higher risk of TBI. It is therefore possible that these factors could confound the associations observed. Finally, people who have experienced TBI are more likely to feel aggression and experience mood disorders after their injury (Azouvi et al., 2017; Ng & Lee, 2019; Ponsford et al., 2018). These have also been linked to changes in emotion recognition and could influence performance on emotion recognition tasks.

1.4.2.1 General cognitive ability

The most common cognitive deficits after TBI are in the domains of memory, working memory, processing speed/attention, and executive functioning skills such as inhibition and planning (Azouvi et al., 2017; Jourdan et al., 2016). A meta-analysis of processing speed and attention difficulties after severe TBI showed a large effect size for deficits in processing speed and attention across tasks (Mathias & Wheaton, 2007). The authors highlighted that slowed processing speed could be having a big impact on the

performance of any time limited tasks used in this population. A meta-analysis of working memory difficulties after moderate to severe TBI reported a small to moderate effect size across visuospatial and verbal tasks (Dunning et al., 2016). Longitudinal studies have shown that the cognitive difficulties after moderate to severe TBI persist long term (Draper & Ponsford, 2008; Jourdan et al., 2016; Ponsford et al., 2014). There is mixed evidence about long term deficits in cognition after mild TBI, with some studies reporting deficits in working memory and processing speed 5 years post injury and others finding no deficits 3 months post injury (Calvillo & Irimia, 2020).

Given the general cognitive deficits prevalent after TBI, it is possible that deficits in emotion recognition performance after TBI are the result of general cognitive deficits. Yim and colleagues (2013) found that verbal and non-verbal memory, working memory performance, and processing speed were all associated with performance on a static face emotion recognition task after severe TBI. They also reported that only non-verbal memory and working memory were predictive of performance on the emotion recognition task. Notably they increased presentation time of stimuli to 15 seconds to try and mitigate the potential impact of processing speed (Azouvi et al., 2017). Conversely, Spikman and colleagues (2012) did not find any evidence for a correlation between tasks measuring general cognitive ability and performance on a facial emotion recognition task. Rosenberg and colleagues (2015) investigated the impact of non-verbal reasoning, working memory, and processing speed on emotion recognition after moderate to severe TBI. They found associations between all three general cognitive domains and emotion recognition performances in the TBI group, but their analysis showed that these did not fully account for emotion recognition deficits observed after TBI. Neumann, Völker, and colleagues (2021) recently investigated the interaction between lesions in brain areas associated with emotion recognition and working memory load on emotion recognition in self and others. They found that both brain lesions and working memory load are associated with decreased emotion recognition performance, but contrary to their prediction, the effects seem to be independent of one another. This provides further evidence that deficits in emotion recognition after TBI are not the result of deficits in general cognitive ability. Nevertheless, it is important to consider how task demands of emotion recognition tasks could be impacting performance given the prevalence of general cognitive difficulties after TBI. Due to the short presentation time of stimuli on the Bristol ERT it is possible that processing speed may be particularly important for performance on the Bristol ERT.

1.4.2.2 Face perception

Arguably, emotion recognition from faces and face identification are two distinct processes that function independently of one another (Duchaine et al., 2003). Evidence for this is provided by the fact that people can be impaired on one but not the other, however, it is likely that there are shared neural mechanisms that underlie both these processes (Haxby & Gobbini, 2011). A functional MRI study by Neumann and colleagues (2014) reported that decreased neural activation of the right fusiform gyrus – associated with face perception was correlated with impaired facial emotion recognition after TBI. They subsequently proposed that difficulties with emotion recognition from faces are driven by general face processing difficulties and not emotion specific problems. It is important to consider that performance on facial emotion recognition tasks is influenced by the ability to perceive faces in general. Consequently, when evaluating emotion recognition in a clinical setting it would be advisable to also consider general face perception and include a measure such as the Glasgow Face Matching Test (Burton et al., 2010; J. D. Henry et al., 2015).

1.4.2.3 Ability to identify emotions (Alexithymia)

Alexithymia refers to difficulties in identifying and understanding one's own emotions, which is a commonly reported difficulty after TBI (Henry et al., 2006). In a neurologically healthy population alexithymia has been linked to deficits in emotion recognition and changes in automatic perception of social cues (Martinez-Sanchez et al., 2017; Parker et al., 1993; Prkachin et al., 2009; Rosenberg et al., 2020). Given the potential links between emotion recognition and expression (Adolphs, 2002b), it is important to consider whether alexithymia could be impacting emotion recognition deficits observed after TBI. In fact, Williams and Wood (2010) suggested that alexithymia could be the cause for emotion recognition deficits after TBI and more recently the ability to experience emotions has been linked to accurate emotion perception in a TBI population (Wearne et al., 2019). It is important to note that the evidence for an association between emotion recognition and alexithymia has been mixed. For example, Rosenberg and colleagues (2019) did not find an association between alexithymia and the Emotion Evaluation Task from TASIT, although alexithymia was correlated with performance on other emotion recognition tasks in that study. There is evidence to suggest that changes in emotion recognition associated with alexithymia are linked to temporal constraints (Ihme, Sacher, Lichev, Rosenberg, Kugel, Rufer, Grabe, Pampel,

Lepsien, Kersting, Villringer, Lane, et al., 2014; Rosenberg et al., 2020). This means that, unlike for the Emotion Evaluation Task from TASIT, alexithymia could impact performance on the Bristol ERT because stimuli are only presented briefly. In sum, it is important to consider the potential impact of alexithymia on emotion recognition in a TBI population.

1.4.2.4 Age

An important consideration is that emotion recognition ability has been shown to decrease with increasing age (Abbruzzese et al., 2019; Ferreira & Torro-Alves, 2016; Mill et al., 2009; Ruffman et al., 2008; West et al., 2012). A study by Byom and colleagues (2019) investigated the interaction between TBI and age on emotion recognition. They found that both having a TBI and older age were associated with decreased emotion recognition but did not find evidence for an interaction between these two factors. The observed decline in emotion recognition with increased age could be the result of general cognitive decline or high comorbidity with mood disorders in an older population (Murphy et al., 2019). Although Horning and colleagues (2012) argued that age related decline in emotion recognition is influenced but not fully explained by decline in general cognitive ability. A recent study by Hayes and colleagues (2020) highlighted the fact that different emotion recognition tasks were associated with different age related effects, indicating that it is important to consider task characteristics when making inferences about emotion recognition.

Another consideration is evidence to suggest that ability to recognise emotions emerges at a young age but continues to develop into late puberty (Lawrence et al., 2015). Some emotions, such as happiness, sadness, and anger are accurately recognised from an early age, whilst recognition for fear, disgust, and surprise increase in accuracy into late puberty (Lawrence et al., 2015). Wu and colleagues (2016) suggest that connectivity between the amygdala and frontal regions of the brain might underly some of these developments, because functional MRI shows differences in activation from childhood through to early adulthood in those areas. Considering TBI in the context of these developments, a TBI that occurred during childhood may not have the same impact on emotion recognition as an injury received after emotion recognition is fully developed around the age of 16.

1.4.2.5 Sex

In line with the popular belief that females are better at recognising emotions than males, studies have shown that females perform slightly better at emotion recognition tasks than males (Thompson & Voyer, 2014; Turkstra et al., 2020). However, the size of this effect is commonly small, and the exact nature of the differences is unclear. It is possible that females have an advantage over men at identifying subtle emotions but not full intensity stimuli (Hoffmann et al., 2010; Sasson et al., 2010). Differences in performance could be specific to negative emotions, but the evidence is mixed and may be confounded by age. For example, females are possibly better at identifying anger (Abbruzzese et al., 2019; Thompson & Voyer, 2014) and potentially have a response bias for fearful and sad faces (Sasson et al., 2010). Like the age effects, the type of emotion recognition tasks used seems to impact results. Abbruzzese and colleagues (2019) reported that older females may be better at discriminating between emotions than older men, but not necessarily better at identifying them.

Although men are more likely to experience a TBI (Headway, 2018; Majdan et al., 2016) it seems that females may be at greater risk of poorer outcomes after mild TBI (REF). A recent study investigating sex differences in outcome after TBI in a European cohort presented evidence for worse outcomes in women after mild TBI but found no sex differences in outcomes reported after moderate to severe TBI (Mikolic et al., 2021). Interestingly, being female may be a protective factor in terms of emotion recognition performance after TBI. Rigon and colleagues (2016) reported that only male participants had decreased emotion recognition on a dynamic emotion recognition task after moderate to severe TBI. However, Zupan and colleagues (2017) were not able to replicate this finding using static facial stimuli. However, their findings did indicate that even in a severe TBI population females have an advantage in recognising fearful expressions compared to males. However, despite the influence of sex being relatively consistent across studies, it is important to emphasise that effect sizes are small and further research is needed to clarify sex and TBI interactions.

1.4.2.6 Aggression

Post TBI aggression is used to describe verbal outbursts as well as physical violence, with the former being more common (Rao et al., 2009; Roy et al., 2017; Sabaz et al., 2014). Aggression is a common sequalae of TBI with estimates of prevalence between 11% and 34% and often negatively impacts rehabilitation and social

relationships (Roy et al., 2017; Tateno et al., 2003). Onset of aggressive behaviours has been observed within minutes of the injury and can persist for years (Ng & Lee, 2019). Roy and colleagues (2017) assessed aggression at three timepoints within a year post injury in a mixed sample of mild, moderate and severe TBI. Scores on the Glasgow Coma Scale were not associated with presence of aggression, indicating that post TBI aggression is not linked to severity of injury and could be prevalent after mild TBI. Aggression after TBI is linked to worse social outcomes and has also been associated with increased alexithymia (Hicks et al., 2017; Roy et al., 2017; Sabaz et al., 2014; Williams et al., 2018).

To the best of our knowledge aggression has not been directly linked to deficits in emotion recognition, although Hoaken and colleagues (2007) found that violent offenders were worse at emotion recognition than non-violent offenders. Further, there is substantial evidence that aggression is associated with hostile attribution, meaning that people with higher levels of aggression are more likely to interpret ambiguous stimuli as hostile (Dodge, 2006). Penton-Voak and colleagues (2013) found that modifying perceptual threshold of ambiguous stimuli towards more positive attribution decreased aggression in young offenders and healthy control participants. Similarly, interpretive bias towards hostility in ambiguous scenarios has been associated with aggression after TBI (D. Neumann et al., 2021). Notably these negative attributions were not correlated with performance on the Emotion Evaluation Test from TASIT. The associations between aggression and performance on emotion recognition tasks are not yet well understood but it is possible that perceptual biases could be linked to adverse behaviours after TBI.

1.4.2.7 Depression and Anxiety

TBI is associated with an increased risk of developing psychiatric disorders such as depression, anxiety, and Post Traumatic Stress Disorder (PTSD; Alway et al., 2016; Ashman et al., 2004; Gould et al., 2011). Prospective studies have shown that pre-injury disorders predispose but do not fully account for the development of disorders post injury (Alway et al., 2016; Gould et al., 2011). Further, the risk of developing a psychiatric disorder seems to be highest in the year following injury (Alway et al., 2016). Major Depressive Disorder, or depressive disorder not otherwise specified are most common after TBI (Alway et al., 2016; Jorge et al., 2004; Ponsford et al., 2018). A meta-analysis examining prevalence of depression indicated an estimated prevalence of

clinically significant levels of depression after TBI in around 38% of people (Osborn et al., 2014). There have been mixed reports of prevalence of anxiety after TBI, in part due to mixed methodologies and definitions of anxiety (Osborn et al., 2016; Ponsford et al., 2018). Nevertheless, a meta-analysis taking into consideration the different approaches indicated that there is increased risk of anxiety after moderate to severe TBI and mild TBI (Osborn et al., 2016). The authors also reported that a diagnosis of Generalised Anxiety Disorder was less prevalent than clinically significant levels of anxiety without diagnosis of an anxiety disorder, estimated to be 11% and 37% prevalence respectively. This highlights the importance of considering continuous measures of anxiety in a TBI population instead of focusing on people with a diagnosis of anxiety. Compared to other mood disorders, anxiety is reported relatively soon after injury and prevalence seems to decrease after the first two years (Alway et al., 2016; Osborn et al., 2016; Ponsford et al., 2018).

There is evidence that both depression and anxiety are associated with deficits in emotion recognition. Recent meta-analytic studies indicate that the deficit in emotion recognition associated with depression is relatively small, reporting Hedge's *g* between -0.16 and -0.22 (Dalili et al., 2015; Krause et al., 2021). Both studies also suggested that recognition of sad faces may not be impaired in people with depression. Demenescu and colleagues (2010) conducted a meta-analysis for emotion recognition after both depression and anxiety. They reported a much larger effect size for depression than the more recent studies (d =-0.58). The estimated effect size for emotion recognition deficits associated with anxiety was d =-0.35. A limitation of the anxiety estimate in their study is that they did not separate findings by type of anxiety disorder included. Plana and colleagues (2014) conducted a meta-analysis looking at individual anxiety disorders and found a small effect size for emotion recognition deficits associated with social phobia, generalised anxiety disorder, obsessive compulsive disorder and panic disorder, but a large effect for deficits after PTSD.

There is mixed evidence regarding the association between emotion recognition and anxiety in sub-clinical populations. It is possible that trait anxiety is associated with increased emotion recognition performance (Attwood et al., 2017; Mendes Ferrer Rosa et al., 2017), with some studies reporting an improvement specifically for fearful stimuli (Doty et al., 2013; Surcinelli et al., 2006). However, other studies have not found any evidence for an association between trait anxiety and emotion recognition (Cooper et al., 2008; Dyer et al., 2022; Suslow et al., 2019). Suslow and colleagues (2019) reported

that state anxiety was also not associated with emotion recognition performance in their study, whilst Attwood and colleagues (2017) reported a deficit for overall emotion recognition after state anxiety was induced. Using the same methodology, Dyer and colleagues (2022) were recently able to replicate this finding and reported that observed deficit was likely driven by decreased emotion recognition accuracy for happiness, disgust and fear. Both trait and state anxiety could also be associated with negative biases in emotion recognition, meaning that ambiguous or neutral stimuli are interpreted as more negative or threating (Attwood et al., 2017; Maoz et al., 2016; Mendes Ferrer Rosa et al., 2017; Richards et al., 2002; Rossignol et al., 2005). Similarly, depression has been associated with a negative attribution bias towards perceiving faces as sad or disgusted (Penton-Voak et al., 2012; Watters & Williams, 2011).

In sum, depression and anxiety are commonly reported after TBI and have been associated with both deficits in emotion recognition and negative attribution biases. Although the reported effects are small it is possible that emotion recognition performance after TBI could be confounded by anxiety and depression, especially after mild TBI (Calvillo & Irimia, 2020). Presence of negative interpretive biases could help understand whether changes in emotion recognition after TBI are being impacted by anxiety and depression. Venkatatesan and colleagues (2021) recently investigated association between social cognition, depression and functional outcomes after TBI. They concluded that social cognition and depression are distinct but interrelated concepts that should be considered conjointly after TBI. To the best of our knowledge research investigating association between emotion recognition and anxiety after TBI has not been published.

1.5 Thesis Overview

Traumatic Brain Injury (TBI) is an injury to the head by an external force that results in changes in brain function and or structure (Menon et al., 2010). Emotion recognition deficits after moderate to severe TBI are well established in the research literature (Babbage et al., 2011; Murphy et al., 2021). However, emotion recognition tasks are not routinely used in a clinical setting (Kelly et al., 2017). It is therefore important to start translating research into clinical practice (Yeates et al., 2017). Kessel (2019) has outlined some of the challenges associated with adapting cognitive tasks developed for research purposes to be suitable for use neuropsychological assessment in clinical practice. One challenge is that the psychometric properties of emotion recognition tasks are not well understood, another that tasks developed in a research setting are not readily available for use in clinical practice. Further, the associations between emotion recognition and other neuropsychological or psychosocial factors need to be established and considered (Cassel et al., 2019). Current understanding of possible changes in emotion recognition after mild TBI is limited due to a lack of research in this area (Calvillo & Irimia, 2020). Given the number of people with mild TBI that report persistent symptoms and disability (Prince & Bruhns, 2017), it is important to investigate whether deficits in emotion recognition are present after mild TBI as well.

The Bristol Emotion Recognition Task (ERT) is a six forced choice emotion recognition task available as part of the Cambridge Automated Neuropsychological Test Battery (CANTAB) owned by Cambridge Cognition Ltd. The task stimuli and procedure are designed to reduce random variation when measuring emotion recognition accuracy, which means that it could also be useful in assessing interpretive bias by considering misidentifications made. This could be useful when considering changes in emotion recognition after TBI and the potential impact of other factors linked to TBI. A feasibility study conducted in 2015 indicated that the Bristol ERT is suitable for use in a clinical setting and able to detect differences in emotion recognition between patients with acquired brain injury when compared to neurologically healthy controls (Müller-Glodde, 2015). As the Bristol ERT is available as part of the CANTAB it could easily become available as a measure of emotion recognition in clinical practice (Kessels, 2019). However, the task has not yet been used to investigate changes in emotion recognition after TBI and there is insufficient information about reliability and validity available for it to be recommended for use in a clinical setting.

1.5.1 Aims

There are three main aims of this thesis. The first is to investigate the psychometric properties of the Bristol ERT to help assess whether the task should be used to measure emotion recognition in a clinical setting. This will be done by checking reliability of the measurements made using the Bristol ERT and considering the validity of inferences made about emotion recognition using the Bristol ERT. The second aim of this thesis is to develop our understanding of changes in emotion recognition after TBI using the Bristol ERT. This will be achieved by trying to replicate the deficit in emotion recognition associated with moderate to severe TBI and investigating the association between emotion recognition and mild TBI to address the gap in the current literature. The third aim is to evaluate whether changes in emotion recognition after TBI could be associated with anxiety. The Bristol ERT has been used to investigate emotion recognition following state anxiety and the results indicated that anxiety can lead to overall emotion recognition deficits and negative attribution bias (Attwood et al., 2017). It follows that anxiety could be influencing emotion recognition performance after TBI and to the best of our knowledge studies have not investigated this association. Ultimately this thesis will add to our understanding of emotion recognition after TBI and help evaluate the utility of the Bristol ERT for use in both clinical and research settings.

1.5.2 Chapter outline

The next two chapters of this thesis will focus on establishing psychometric properties of the Bristol ERT in neurologically healthy populations to gain a better understanding about whether reliable and valid inferences about emotion recognition can be made based on performance on the Bristol ERT. In Chapter 2, the test-retest reliability of the Bristol ERT will be investigated in a longitudinal cohort study and a construct validation study comparing performance on the Bristol ERT to performance on two other emotion recognition tasks is presented in Chapter 3. The study presented in Chapter 4 investigates emotion recognition deficits using the Bristol ERT in a clinical setting by comparing performance between patients with moderate to severe TBI to age and sex matched healthy controls Chapters 5 and 6 will look at associations between emotion recognition, mild TBI, and anxiety. Chapter 5 presents an online observational study comparing performance on the Bristol ERT between participants with self-reported TBI and neurologically healthy controls In Chapter 6, data from a prospective longitudinal cohort study is analysed to explore these associations.

Chapter 2 Test-retest Reliability of the Bristol Emotion Recognition Task in the Avon Longitudinal Study of Parents and Children

2.1 Introduction

To effectively use and interpret the results of any given test it is necessary to understand the psychometric properties of the test (Bowden, 2017). Consequently, standardised tests used in clinical practice are required to establish information about reliability and validity (Mitrushina et al., 2005; Strauss et al., 2006). This is because outcomes are used to inform clinical decision making and scores that unreliable or do not allow for valid inferences could lead to incorrect diagnoses or treatment. Although tasks used in research settings are not always held to the same standard as in clinical practice it is equally important for theory to be developed based on reliable data and valid inferences. There is a growing awareness in the field of cognitive research that there is a lack of understanding about the psychometric properties associated with commonly used cognitive measures (Parsons et al., 2019). In 2016, the National Advisory Mental Health Council Workgroup on Tasks and Measures for Research Domain Criteria highlighted the lack of psychometric data for tasks used in research and recommended both further research and a standardisation of practices in the field (NAMHC, 2016). Despite a need for research around the development of cognitive tasks, studies addressing this topic are often harder to fund and publish (Vitoratou & Pickles, 2017).

To date little information has been published about the reliability and validity of the Bristol Emotion Recognition Task (ERT), the primary measure of emotion recognition used in this thesis. To evaluate general utility of this task for research or in a clinical setting these psychometric properties need to be better established. Reliability is a requirement, though not sufficient, to infer validity (Cook & Beckman, 2006), so the first logical step is to investigate the reliability of the Bristol ERT. This chapter will explore what reliability means, review relevant literature regarding reliability of the Bristol ERT, and then assess test-retest reliability of the Bristol ERT using data collected at as part of a longitudinal cohort, the Avon Longitudinal Study of Parents and Children (ALSPAC). At the same time as the Bristol ERT, participants in this cohort also completed two commonly used measures of cognitive ability, a working memory task (N-back) and a measure of inhibition (Stop Signal Task). Calculating reliability coefficients for these tasks using the ALSPAC data and comparing these to reliability estimates reported in the literature will help establish psychometric information for the Bristol ERT.

2.1.1 What is reliability?

Broadly, a task is considered reliable if measurements made using that task can be replicated given stable conditions (Bowden, 2017; Streiner et al., 2008). According to Classic Test Theory every time something is measured there will be an error in measurement, so the obtained score is a combination of an individual's true score and an error in measurement (DeVellis, 2006; Novick, 1966). Good reliability is obtained if the error in measurement is small, meaning that the observed score is close to the true score. The reliability coefficient is expressed as the ratio of the variability in the true score to variability in the obtained score. As the true score for an individual is unknown, it is not possible to determine what proportion of the observed score is the result of error. Instead, we estimate reliability by using variance (σ^2) in measurement as an indicator of variability.

$$Reliability = \frac{Subject\ variability}{Subject\ variability + Measurement\ error} = \frac{\sigma^2_{subject}}{\sigma^2_{subject} + \sigma^2_{error}}$$

(Fisher, 1925; Streiner et al., 2015)

There are different types of reliability that can be estimated using variance. They consider different sources of error in measurement and are in turn calculated in different ways (Streiner et al., 2015). Internal consistency assesses whether items within a test measure the same construct. Commonly reported reliability coefficients are Cronbach's alpha, and a split half estimate based on a correlation using Spearman-Brown correction (Hogan et al., 2000; Reis & Judd, 2000; Streiner et al., 2008). Test-retest reliability evaluates the stability of measurement over time, parallel forms reliability assesses the equivalence across versions of the same test, and inter-rater reliability assesses variability in judgment made by different people. Commonly reported reliability coefficients for these types of reliability are Pearson's-product-moment correlation or an Intraclass Correlation Coefficient (Reis & Judd, 2000; Streiner et al., 2008; Weir, 2005). All these coefficients use a correlational approach to estimate reliability and the most appropriate reliability coefficient will depend on the data collected and question being asked of that data. For example, internal consistency can be assessed when there is only a single point of measurement, whilst the other types of reliability require two or more measurements.

It is important to note whilst tasks are often described as reliable or unreliable, they are not inherently one or the other (Hedge et al., 2018; Parsons et al., 2019). It is the measurements made using a task that can be assessed for reliability. This means that reliability coefficients for a task are calculated for a specific sample, and the reliability estimates can vary across samples because every time a task is used the error in measurement will be slightly different. Therefore, using the reliability calculated for one sample to infer reliability of measurement in a different sample has been described as reliability induction (Parsons et al., 2019; Vacha-Haase et al., 2002). Reliability induction is not ideal but is often inevitable as reliability estimates are not available for every sample or possible to calculate for individual measurements. Parsons and colleagues (2019) recommend that, in a research setting, reliability estimates should be calculated and reported for every study and reliability induction should be avoided where possible. For research in novel populations pooled estimates from previous studies could be used as a benchmark for comparable reliability instead of applying arbitrary thresholds to evaluate reliability. These pooled estimates are also critical in a clinical setting, where tasks are commonly used as part of an assessment to evaluate individual performance (Bowden, 2017). Well established reliability estimates based on normative data or studies that have assessed reliability in a population of interest are required to accurately interpret individual test scores. They help determine the likely true range of an individual score and whether a change in performance is likely to reflect a true change in cognitive ability (Vaz et al., 2013).

Finally, reliability coefficients are not the only way in which reliability is evaluated. In an experimental setting, tasks are often considered reliable if they repeatedly replicate a given effect with a consistent effect size, but these tasks are not necessarily associated with good reliability coefficients (De Schryver et al., 2016; Hedge et al., 2018; Soveri et al., 2018). In experimental research the aim is often to minimise individual differences within a sample in order to effectively compare group effects (De Schryver et al., 2016). Hedge and colleagues (2018) argue that the resulting homogeneity within these groups causes the individual scores to be close together when ranked. Consequently, if the same individuals complete the chosen measure again, they are less likely to be ranked consistently. This results in low reliability estimates based on a correlational approach commonly used to calculate reliability coefficients. It does not however mean that these tasks are inappropriate to investigate group differences (Hedge et al., 2018). For example, the Bristol ERT may reliably differentiate between a

group of participants with TBI and a group of controls even if individual variance in performance is high (i.e., reliability coefficients are low). However, if a task is being used to assess individual differences it is crucial to use tasks that have been shown to have high reliability estimates or consider potential implications when this is not possible.

In sum, for the purpose of this study, reliability is assessed using reliability coefficients to evaluate the utility of the tasks included to assess change in performance of an individual. This means that reliability is assessed through individual variability in measurement. Parsons, Kruijt, and Fox (2019) propose standardised practices for reporting reliability in cognitive behavioural tasks. Split-half reliability should be calculated to assess internal consistency when there is data from only one measurement, whilst test-retest reliability should be calculated if data is available from two or more timepoints. In ALSPAC data for the Bristol ERT and the two other cognitive tasks is available at two timepoints. This study will investigate test-retest reliability of these tasks as a way to assess to the stability of measurement across time.

2.1.2 Reliability coefficients for test-retest reliability

Pearson's correlation coefficient has commonly been used as a reliability estimate for test-retest reliability, but more recently the Intraclass Correlation Coefficient (ICC) has been recommended as a more appropriate indicator (Hedge et al., 2018; Koo & Li, 2016; Parsons et al., 2019; Polit, 2014; Weir, 2005). An important difference between the two is that the ICC can account for systematic error in the data, whilst Pearson's r only captures consistency, i.e., whether individuals are ranked in the same order at the two testing points (Berchtold, 2016; Bruton et al., 2000; Koo & Li, 2016; Weir, 2005). Koo and Li (2016) argue that when assessing reliability it is important to measure absolute agreement between scores, and not consistency in raking, by accounting for systematic changes. There are six versions of the ICC presented by Shrout and Fleiss (1979) and ten versions according to McGraw and Wong (1996). They are all variations of the reliability formula presented in the previous section, in fact, the ICC1.1 according to Shrout and Fleiss (1979) is exactly that:

$$ICC1.1 = \frac{\sigma^2_{subject}}{\sigma^2_{subject} + \sigma^2_{error}} =$$

$$\frac{Mean\ Square\ _{between\ subjects} - Mean\ Square\ _{within\ subjects}}{Mean\ Square\ _{between\ subjects} + (number\ of\ trials - 1)\ \times\ Mean\ Square\ _{within\ subjects}}$$

(Shrout & Fleiss, 1979; Weir, 2005)

The type of ICC most appropriate to use depends on whether the data is based on a single or mean score, a 1-way or 2-way model is being investigated, effects are thought to be fixed or random, and whether systematic change should be accounted for (Weir, 2005). Arguably, the most appropriate ICC for test-retest analyses is the ICC2.1 (Shrout & Fleiss, 1979), because it accounts for systematic change by treating it as a form of error. It is used to estimate agreement between scores at two time points (Berchtold, 2016; Koo & Li, 2016; Parsons et al., 2019).

$$ICC2.1 = \frac{Variance\ between\ individuals}{Varaiance\ between\ individuals + Error\ variance + Variance\ between\ sessions}$$

(Hedge et al., 2018)

The ICC3.1 (Shrout & Fleiss, 1979) does not account for variance between sessions and thus is an indicator of consistency as opposed to agreement in scores. It may be the most appropriate reliability coefficient to use if systematic changes, like practice effects, are expected (Parsons et al., 2019). As a measure of consistency, the ICC3.1 is likely to closely match Pearson's r.

Publications often fail to distinguish between consistency and agreement and regularly do not specify what ICC is used when reporting test-retest reliability (Berchtold, 2016). Information about both consistency and agreement can be useful if they are reported clearly and interpreted appropriately. Measures of agreement factor in systematic error, so naturally they are lower if there is a systematic shift in the data. Measures of consistency are less affected by systematic changes and thus may be more appropriate to use if systematic shifts are expected or not considered important. The ICC2.1 will be reported as a measure of agreement in this study. In addition, the ICC3.1 and Pearson's r will be reported as measures of consistency.

2.1.3 Interpreting reliability coefficients

Standardised reliability coefficients for test-retest reliability fall between 0 and 1 because of how they are calculated (Streiner et al., 2008; Webb et al., 2006). For the ICC, the variance between individuals is always going to be less than the variance between individuals plus error variance, and for Pearson's r, a negative coefficient would be implausible when correlating scores on the same task. Higher coefficients are always indicative of better reliability making it easy to compare coefficients across studies and tasks. The difficulty is deciding what level of reliability is considered adequate, which will

vary depending what the task is being used for (Strauss et al., 2006). For example, tasks used for decision making will be required to be associated with much higher reliability coefficients than a task used to assess outcomes. A variety of acceptable thresholds have been suggested, and terms used to describe these thresholds are inconsistent and often overlapping (Cicchetti, 2001; Koo & Li, 2016; Ponterotto & Ruckdeschel, 2007; Strauss et al., 2006). Strauss and colleagues (2006) proposed the following thresholds as general guidelines when evaluating the reliability of neuropsychological test: coefficients less than .60 are low, coefficients between .60 and .69 are marginal, coefficients between .70 and .79 are adequate; coefficients between .80 and .89 are high, and coefficients greater that .90 are very high. With regards to the ICC Koo and Li (2016) suggested that an ICC <0.5 is indicative of poor reliability, an ICC between 0.5 and 0.75 shows moderate reliability, between 0.75 and 0.9 indicates good reliability, and an ICC greater than 0.90 demonstrates excellent reliability. Whilst these thresholds can be useful, it is important to remember that they are arbitrary points on a continuous spectrum and should not be treated as definitive (Parsons et al., 2019). A better approach might be to describe reliability being within a range based on the 95% confidence intervals (Koo & Li, 2016). Parsons and colleagues (2019) proposed avoiding use of thresholds altogether, and instead a normative range of reliability coefficients could be established by pooling reliability estimates for any given task. Arguably, evaluating reliability in the context of previous findings is more informative than applying arbitrary thresholds to determine whether measurements are reliable.

Reliability coefficients are calculated on the assumption that all within subject variance is due to an error in measurement, be it random error or systematic error. When interpreting reliability coefficients, it is important to consider factors other than measurement error that could be causing variance or influencing the reliability coefficient. For instance, as mentioned previously, a homogenous sample can be problematic because when there is very little between subject variance, small changes in measurement error have a bigger impact on the reliability coefficient (Hedge et al., 2018). In the case of test-retest reliability it is also important to consider the impact of the test-retest interval, i.e., time between testing sessions. Participants commonly perform better on a task the second time they are asked to complete it, known as a practice effect (Calamia et al., 2012). If reliability is assessed as agreement between scores (e.g., ICC2.1) this systematic change will result in a lower reliability coefficient, although less likely to impact estimates of consistency (Streiner et al., 2008). Arguably,

practice effects are stronger when there is a shorter test-retest interval, although they have been shown to persist for several years (Calamia et al., 2012). Equally, a long interval between testing sessions could result in a true change of the construct being measured, which would also result in a lower reliability coefficient. Consequently, there is no standard recommended time between testing sessions, although 2 to 14 days is a common interval (Streiner et al., 2008). Another important consideration is that true change in the underlying construct is more likely if it is state dependent as opposed to a trait, in which case test-retest reliability may not be the best indicator of reliability (Polit, 2014). Finally, contextual factors that differ between testing sessions could give rise to variation in measurement that impacts the reliability estimates (Webb et al., 2006). These considerations are important when evaluating reliability coefficients, as well as informing study design or the way in which tasks are used.

Test-retest reliability is useful in a clinical setting because tasks are often administered at different times to see whether there has been a change in performance for an individual (Lexell & Downham, 2005). It is sometimes difficult to apply standardised reliability coefficients such as the ICC or Pearson's r to the interpretation of individual scores (Streiner et al., 2008), and it is often more useful to express reliability in the units of the task used to make that measurement. The Standard Error of Measurement (SEM) and the Smallest Real Difference (SRD) are absolute indicators of reliability that are useful when interpreting individual scores (Lexell & Downham, 2005; Vaz et al., 2013; Weir, 2005). The SEM is an estimate of the within-subject standard deviation and can be expressed as:

$$SEM = \sqrt{within\ subject\ variability} = standard\ deviation\sqrt{1 - reliability\ coefficient}$$

(Streiner et al., 2008; Weir, 2005)

Whilst the SEM is related to the ICC, it provides information about the precision of individual scores instead of quantifying how well scores obtained from different people are differentiated (Streiner et al., 2008). The SRD, also known as repeatability coefficient or minimal difference, can be defined as the minimum difference between scores needed to indicate change beyond random error (Vaz et al., 2013; Weir, 2005). The SRD is based on the 95% confidence interval around the SEM and can be expressed as:

$$SRD = \left(1.96\ \times \sqrt{2}\ \right) \times SEM = 2.77\ \times SEM$$

These absolute indicators of reliability are helpful when interpreting individual scores, but it is important to note that they do not show whether the change in test score observed translates to meaningful change in functioning (Vaz et al., 2013).

2.1.4 Why is it important to establish reliability?

Reliability impacts the ability to detect true change and make accurate inferences based on our data. In a clinical context, tasks that are associated with low reliability coefficients and in turn a high standard error of measurement, are less likely to provide useful information about individual performance (Bowden, 2017; Vaz et al., 2013). Reliability can also influence statistical power and should be taken into consideration when conducting sample size calculations, and when comparing effect sizes across studies or tasks (Parsons et al., 2019). Low statistical power has been highlighted as a pervasive issue in psychological research (Anvari & Lakens, 2018; Button et al., 2013). It is associated with an increased likelihood of making a type II error (i.e., claiming there is no effect when there is in fact a true effect), as well as decreasing the positive predictive power (i.e., decreasing the likelihood of correctly identifying a true effect). The relationship between reliability and statistical power is not straight forward, as demonstrated mathematically by Zimmerman and Zumbo (2015). They show that increasing reliability will only increase statistical power if the increase in reliability is due to a decrease in error variance and not an increase in subject variance. Parsons and colleagues (2019) argue that it is beneficial to consider the impact that reliability could be having on statistical power when interpreting results.

Conducting a-priori sample size calculations to indicate the number of participants needed in a study to detect an effect size of interest with adequate power at a set significance criterion, can help avoid running underpowered studies (Lakens, 2013b). Determining an accurate effect size estimate is crucial to ensure that enough data is collected for us to be confident in the statistical inferences made based on that data. When using correlation coefficients as a measure of effect size it is important to consider that the reliability of tests limits the maximum observable correlation between variables (Parsons et al., 2019). Spearman's (1904) attenuation correction formula states:

$$True\ correlation = \frac{Observed\ correlation}{\sqrt{(reliability\ of\ task\ 1\ \times reliability\ of\ task\ 2)}}$$

It would be good to consider the impact of reliability could be having on an observed effect size and where possible adjust effect size estimates before doing a sample size calculation to avoid conducting an underpowered study (Parsons et al., 2019). For example, if the true correlation is estimated as *r* = 0.5 and the reliability associated with both tasks is considered high (*r* = 0.8 and *r*= 0.85) then the maximum observable correlation is *r* = 0.41 (based on Spearman's formula rearranged to calculate the observed correlation $0.5 \times \sqrt{(0.8 \times 0.85)}$). The lower estimate of r= 0.41 should be used as an effect size for the sample size calculation. Parsons and colleagues (2019) also argue that when comparing results across studies the effect sizes in each study should be corrected using the reliability reported for the sample in that study. When reliability is not accounted for, group differences could be the result of differing error variance in the samples being compared (Cooper et al., 2017).

2.1.5 Reliability of the Bristol Emotion Recognition Task (ERT)

Bland and colleagues (2016) conducted a study to validate the EMOTICOM battery (a Neuropsychological Test Battery to Evaluate Emotion, Motivation, Impulsivity, and Social Cognition) which included a version of the Bristol ERT. Faces expressing four of the six available emotions were presented (happy, sad, angry, and fearful). Their test-retest reliability estimates are based on 42 participants asked to repeat the battery between five to ten days after being tested for the first time. They reported good test-retest reliability for the Bristol ERT (ICC2 = 0.86) and concluded that the battery, including the Bristol ERT, is a valid assessment tool. Notably they only reported the reliability estimate for an affective bias score calculated by subtracting the average score for sad faces from the average score for happy faces on the Bristol ERT. Despite listing average accuracy on the Bristol ERT as an outcome for the battery, no reliability estimates for average emotion recognition accuracy was reported in this paper. It is unclear how or why the affective bias score and not average emotion recognition was chosen as the outcome measure for this study, especially as this means performance on angry and fearful faces was not considered at all.

Test-retest reliability of emotion recognition accuracy has been investigated using other emotion recognition tasks. For example, Palmer, Langbehn, Tabrizi, and Papoutsi (2018) recently conducted a test-retest reliability analysis of emotion recognition accuracy using 60 Ekman and Friesen faces for happiness, sadness, anger, fear, disgust, surprise and neutral as stimuli. They tested 16 participants at three

different timepoints, each approximately a week apart. Test-retest reliability for accuracy of negative emotions (combined accuracy for anger, fear, disgust, and sadness) was calculated using a Spearman correlation. They reported high reliability of r = 0.84 for both the comparison between sessions one & two, and sessions two & three. The reason they use only combined accuracy of negative emotions as an outcome measure is because that is what is used to test for striatal impairments in a clinical setting. They did not find an overall practice effect for accuracy. This data could indicate that the Bristol ERT would have relatively good/high test-retest reliability for emotion recognition accuracy. However, it is possible that the reliability scores reported by Palmer and colleagues (2018) were inflated considering they had a relatively small sample and it does not give a reliability estimate for overall emotion recognition accuracy, only accuracy of negative emotions.

A further consideration is that the task demands for the Bristol ERT are likely to be different to the task used by Palmer and colleagues (2018) because emotions in the Bristol ERT are presented at varying levels of intensity. As accuracy is decreased for low intensity stimuli the test-retest reliability of the task is also likely to decrease, although ceiling effects at 100% intensity could also result in low reliability estimates. Evidence for this pattern of results is provided by Cecilione and colleagues (2017) in a study investigating test-retest reliability of a facial expression labelling task similar to the Bristol ERT. Participants in this study were twin children aged 9 to 14 years who attended testing sessions two to five weeks apart. An unbiased hit rate (As described by Wagner, 1993) was used as their main outcome measure and emotion specific reliability, but not the reliability for overall emotion recognition accuracy was reported. The results indicated that reliability was decreased at lower intensity presentations across all emotions and the highest reliability estimates for each emotion varied between 60%-100% intensity. The highest Pearson correlation coefficient for each emotion were: anger at 90% intensity r= 0.508 (0.375, 0.621), fear at 70% intensity r= 0.465 (0.318, 0.592), happiness at 80% intensity r= 0.587 (0.461, 0.691), sadness 100% intensity r= 0.467 (0.323, 0.590), disgust at 70% intensity r= 0.6 (0.480, 0.698), and surprise at 60% intensity r= 0.440 (0.291, 0.568). The fact that the high reliability did not automatically correspond to the highest level of intensity (100%) could indicate that there were ceiling effects at 100% intensity, which resulted in homogeneity and could have decreased reliability estimates. The authors proposed that median scores for each emotion better accounted for the impact of intensity on performance, which they calculated as an

intercept score using latent growth modelling to analyse their data. When using these intercept scores, the reliability coefficients were much better: anger r= 0.811 (95% CI 0.661, 0.829), fear r= 0.796 (95% CI 0.652, 0.897), happiness r= 0.831 (95% CI 0.741, 0.899), sadness r= 0.763 (95% CI 0.619, 0.864), disgust r= 0.847 (95% CI 0.715, 0.934), and surprise r= 0.789 (95% CI 0.606, 0.904). The scores were less impacted by high variability at the lower intensities and modelling the data in this way seems to be an effective method to analyse emotion specific data when trial level data is available. Cecilione and colleagues (2017) concluded that their task showed strong test-retest reliability based on the intercept reliability, despite the correlation coefficients for the unbiased hit rates being low across the board. This highlights the importance of considering what outcome measure are appropriate.

Adams and colleagues (2016) investigated psychometric properties for a range of emotional processing tasks in adults, again including a facial emotion recognition task comparable to the Bristol ERT. They used a discrimination index as their main outcome measure for each emotion and report a two-way mixed model ICC as an estimate of reliability: anger ICC = 0.577 (95% CI, 0.167, 0.798), disgust ICC = 0.984 (95% CI, 0.961, 0.994), fear ICC = 0.791 (95% CI, 0.536, 0.914), happy ICC = 0.689 (95% CI, 0.353, 0.868), sad ICC = 0.840 (95% CI, 0.631, 0.935), and surprise ICC = 0.597 (95% CI, 0.208, 0.823). Using the thresholds proposed by Koo and Li (2016), these scores range from moderate to excellent (Cicchetti, 2001), although some of the confidence intervals were very big and suggest that reliability could be poor. The authors caution that their sample was too small to reach adequate power and that results cannot be generalised to other tasks.

In sum, there is some evidence that suggests test-retest reliability of the Bristol ERT should be adequate, but studies vary in the tasks they use, outcome measures, and analysis techniques. Latent growth modelling seems to be a good approach to capture emotion specific trends, but as trial level data was not available for the current study, this approach was not used here. Instead, test-retest reliability for emotion specific accuracy based on an unbiased hit rate will be assessed. None of the studies discussed have reported test-retest reliability for overall emotion recognition based on positive and negative emotions combined. There is also no discussion of or reasoning for use of consistency estimates versus estimates of agreement and there also seems to be a tendency to rely on small sample sizes when investigating test-retest reliability. The primary aim of this study was to investigate test-retest reliability of overall emotion accuracy (total hits) using the Bristol ERT in a large sample of healthy young adults.

2.1.6 Avon Longitudinal Study of Parents and Children (ALSPAC)

Pregnant women resident in Avon, UK with expected dates of delivery 1st April 1991 to 31st December 1992 were invited to take part in the ALSPAC study. The initial number of pregnancies enrolled is 14,541. Of these initial pregnancies, there was a total of 14,676 foetuses, resulting in 14,062 live births and 13,988 children who were alive at 1 year of age. When the oldest children were approximately 7 years of age, an additional 913 children not included in the initial phase were enrolled. Details about the phases of enrolment and further information about the cohort are described in the cohort profile papers and update (Boyd et al., 2013; Fraser et al., 2013; Northstone et al., 2019). The total sample size for analyses using any data collected after the age of seven is 15,454 pregnancies, resulting in 15,589 foetuses. Of these 14,901 were alive at 1 year of age.

Data for the three cognitive tasks reviewed in this chapter was collected at a clinic when the children, now young adults, were around 24 years of age. The clinic was held between June 2015 and October 2017, with 4026 of the young people attending the clinic in that time period. Gaining a better understanding of the reliability associated with these cognitive tasks in this cohort could be useful in the interpretation of research that has already been conducted, as well as to inform future research. For example, a study by Mahedy and colleagues (2020) used these cognitive tasks as an outcome measure for cognitive functioning following tobacco and alcohol use. Additionally, comparing these reliability estimates for this sample to those reported in the literature will potentially allow for an evaluation of the quality of data collected in this cohort. If reliability estimates are low across all three tasks it could indicate poor-quality data.

2.1.7 Aims and Hypotheses

In this chapter, test-retest reliability of overall emotion recognition accuracy (total hits) measured using the Bristol ERT was estimated in a large cohort study. Based on the current literature we expected adequate to high reliability of the Bristol ERT in this sample according to the thresholds defined by Strauss and colleagues (2006). Reliability estimates for the N-back and Stop Signal Task will also be reported for the purpose of evaluating the reliability of the measures in the ALSPAC cohort (for a review of reliability estimates for these tasks see Appendix A, section 8.1). Data from these tasks is available to all ALSPAC researchers, so they are likely to be used to address a variety of research questions. Consequently, reliability estimates of both consistency and agreement of scores across the two testing sessions at age 24 will be reported.

2.2 Methods

2.2.1 Participants

The participants were drawn from the Avon Longitudinal Study of Parents and Children (ALSPAC). Around age 24, participants enrolled in ALSPAC were invited to attend a clinic-based assessment. Of those who attended the clinic, 3854 provided valid data for the cognitive tasks assessed in this study. Approximately 3% of participants who attended that clinic were invited to return 4 to 10 weeks later to complete the same tasks again. 112 participants attended both timepoints and cognitive task data for both the first and second timepoint were extracted for these participants. Ethical approval for the study was obtained from the ALSPAC Ethics and Law Committee and the Local Research Ethics Committees.

2.2.2 Measures

The measures of interest were three computer-based cognitive tasks. The Bristol ERT, a visioverbal version of the N-back task (Kirchner, 1958), and the Stop Signal Task (Logan & Cowan, 1984), used to assess emotion recognition, working memory and inhibition respectively. All three tasks were presented using E-prime software (PST Inc, Sharpsburg, PA, USA). Study data were collected and managed using REDCap electronic data capture tools hosted at the University of Bristol. REDCap (Research Electronic Data Capture) is a secure, web-based software platform designed to support data capture for research studies(Harris et al., 2009). Please note that the study website contains details of all the data that is available through a fully searchable data dictionary and variable search tool (http://www.bris.ac.uk/alspac/researchers/data-access/data-dictionary/).

2.2.2.1 Emotion recognition

The Bristol ERT was used to assess emotion recognition. Please refer to the introductory chapter (section 1.3.2) for a full description of the task. The only deviation from that description is that stimuli were presented for 200ms instead of 500ms and the response screen timed out after 10 seconds and automatically started the next trial. The primary outcome measure for the current study was the proportion of total hits on the Bristol ERT calculated by dividing the total number of hits by 96 (total number of trials). The secondary outcome was individual emotion recognition accuracy based on the unbiased hit rate (Hu) as outlined by Wagner (1993). For both measures, higher scores are indicative of better performance on the Bristol ERT.

2.2.2.2 Working memory

The N-back task (2-back condition) was used as a measure of working memory. Participants were presented numbers between 0 and 9 on a screen. For each presentation they had to indicate whether the number in the current trial was the same or not as the number presented 2 trials before. The numbers were presented in black on a white background for 500ms, after which participants were presented with a response screen for 3000ms. After completing a practice block with 12 trials, participants completed an experimental block of 48 trials. The experimental block included eight targets, which are the trials where the number in the current trial is the same as the number presented two trials before. The outcome measures for this task are the number of times a participant correctly identified a target (hits) and the number of times a participant erroneously thought there was a target (false alarms). If a participant pressed 'same' every trial they would have a perfect hit score but also high false alarms. Hence, the hits and false alarms are used to calculate a discriminability index (*d*'prime) as a measure of overall performance:

$$d'prime = inverse\ normal\ (hits) - inverse\ normal\ (false\ alarms)$$

(Rossi et al., 2016)

The *d*'prime was used as the primary outcome for the N-back in this study. A positive *d*'prime indicates good discrimination of hits versus false alarms, whilst a negative *d*'prime indicates low discriminability.

2.2.2.3 Inhibition

The Stop Signal Task was used to assess response inhibition. During the task participants are presented with an X or an O on the screen and asked to press a corresponding response button identifying the letter as quickly as possible. In 25% of trials the participants hear a beep shortly after the presentation of the letter and are instructed to not give a response if they hear the beep i.e., the stop signal. Each trial started with the presentation of a fixation cross in the middle of the screen and ends once the participant has made a response. Participants completed 32 practice trials and then 4 blocks of 64 test trials i.e., 256 test trials in total. Each block was split into sets of 16 trials that each contained 12 trials without a stop signal (go trials) and 4 trials with a stop signal (stop trials) presented in a random order. Both accuracy and reaction time for each trial was recorded. The primary outcome was a Stop Signal Reaction Time (SSRT) which was calculated based on median SSRT proposed by Band, van der Molen

and Logan (2003). It is calculated using the reaction time on the go trials and the delay between the letter appearing and the stop signal being presented:

$$SSRT_{med} = difference\ between\ the\ median\ inhibition\ function\ and\ median\ reaction\ time$$

Shorter times on the SSRT are indicative of better inhibition as theoretically less time is needed to inhibit a response after a stop signal is given.

2.2.3 Data extraction and withdrawal

Variables extracted from ALSPAC included: the raw data required to calculate the primary outcomes for each cognitive task, demographic information (age at clinic, sex, and current occupation), as well as information about length of time between testing sessions. Data was extracted using Stata on the 19th October 2019 and cohort withdrawal scripts were applied to exclude any participants that have requested their data be withdrawn from ALSPAC at that time. The final data set was exported into RStudio (2020) R version 4.0.4 for data cleaning and analysis. The main analysis packages used were the core *stats* package in R, *psych* package version 2.1.9, and *Hmisc* package version 4.4.2 (Harrell, 2014).

2.2.4 Data Analysis

The main analysis was a calculation of the reliability coefficients for the primary outcomes on the Bristol ERT, N-back, and SST as an indication of test-retest reliability of these tasks in ALSPAC. The primary outcome measure for each task were proportion of total hits for the Bristol ERT, *d*'prime for the N-back, and $SSRT_{med}$ for the SST at the both testing sessions. As recommend by Parsons and colleagues (2019), the ICC2.1 was reported as a measure of agreement between scores between the testing sessions and the ICC3.1 was reported as a measure of consistency. Additionally, Pearson's r based on the correlation between testing session 1 and testing session 2 was reported for each task as this is a commonly reported indicator of reliability in the current literature.

To investigate whether there could be a systematic change in performance between testing sessions (for example due to practise effects), a paired Welch's t-test was conducted. To evaluate the size of this effect, a version of Hedge's g and common language effect size was reported (Lakens, 2013b). The effect size estimate was adjusted using the reliability estimates according to Baugh (2002). Further, the effect of length of time between testing sessions was explored by splitting participants into short or long

delay between sessions groups. The ICC2.1 was used to calculate Standard Error of Measurement (SEM) and Smallest Real Difference (SRD), to indicate the size of change required to indicate a true change in performance for an individual based on the reliability of this sample.

A secondary analysis of the Bristol ERT data was conducted looking at emotion recognition accuracy for each of the six emotions individually. The outcome measure for this analysis was an unbiased hit rate based on the number of hits and false alarms for each emotion. Like the primary analysis, the reliability coefficients for each emotion were calculated and systematic change between sessions investigated using a paired Welch's t-test.

2.3 Results

2.3.1 Demographics

The 112 participants (47 male, 65 female) that attended the cognitive testing session at both time points had a mean age of 24 years (range 22 to 26 years). At the time of the first testing session most participants, 75, reported being in full- or part-time employment. 17 participants reported being in full- or part-time education or training, and 10 participants said they were engaged in both work and education. Only 9 participants indicated that none of those options describe their current occupation. The mean time between the first and the second testing session was 43 days with a range of 27 to 77 days in between sessions. The median was slightly lower at 40.5 days between testing sessions.

2.3.2 Exclusions

For each task, participants were excluded if they had missing data at one of the two testing timepoints. Of the 112 participants that attended both testing session 22 had missing Bristol ERT data, 23 had missing N-back data, and 24 had missing SST data. Additionally, 2 participants were excluded from the N-back analysis as their responses were considered unreliable (defined as having more false alarms than non-target correct responses, most likely indicating confusion of response keys). This means that the final *n* used for the analyses was slightly different for each task: 90 for the Bristol ERT, 87 for the N-back and 88 for the SST. There was substantial overlap of missing data across tasks, meaning that most participants who had missing data for one task also had missing data for the other two tasks. Exclusion of all participants that had any missing data left 85 cases with complete data for all three outcome measures. The analysis was re-run using this complete data set as a sensitivity analysis.

2.3.3 Reliability coefficients for the primary outcomes

Reliability coefficients for each of the three tasks were calculated to assess the stability of measurement between the first and second testing sessions. Results for all the tasks are presented in Table 2-1. For the Bristol ERT both ICC scores suggested moderate reliability in this sample based on the thresholds outlined by Koo and Li (2016). Similarly, Pearson's r indicated marginal reliability according to the thresholds set by Strauss and colleagues (2006). When the confidence intervals around these estimates were taken into consideration, as suggested by Koo and Li (2016), the test-

retest reliability in this sample should be reported as poor to moderate or low to adequate. This means that overall emotion recognition scores were associated with quite a lot of within-subject variability between sessions.

Table 2-1

Test-retest reliability coefficients for the three cognitive tasks

Task (outcome measure)	Intraclass Correlation Coefficient (ICC)					Pearson's r		
	df	**ICC2.1**	**95% CI**	**ICC3.1**	**95% CI**	**df**	**r**	**95%CI**
Bristol ERT (prop. of totalhits)	89	0.56	0.36, 0.70	0.60	0.45, 0.72	88	0.60	0.45, 0.72
N-back (*d*'prime)	86	0.43	0.24, 0.59	0.43	0.24, 0.59	85	0.43	0.24, 0.58
SST ($SSRT_{med}$)	87	0.62	0.30, 0.79	0.70	0.58, 0.80	86	0.73	0.62, 0.82

Note: df is degrees of freedom. CI refers to 95% Confidence Interval. ERT is Emotion Recognition Task. Prop. is proportion. SST is Stop Signal Task

The reliability coefficients for the N-back task indicated poor/low reliability of measures in this sample, with the confidence intervals in a poor to moderate range. The reliability coefficients for the SST were slightly higher and indicated moderate/adequate reliability. The confidence intervals for the ICC3.1 and Pearson's r were in the moderate/marginal to good/high range, but the confidence interval for the ICC2.1 was very large, suggesting that reliability could be poor/low in this sample. As expected for all three tasks the two measures of consistency, ICC3.1 and Pearson's r were very similar. For the Bristol ERT and SST, the ICC2.1 was slightly lower than the ICC3.1 and Pearson's r. This means that the agreement between scores was not as good as the consistency (i.e., ranking of participant scores), suggesting that there was a systematic shift in performance between the two testing sessions indicative of a practice effect.

2.3.4 Comparison of means for the primary outcomes

The descriptive statistics for the primary outcomes on the Bristol ERT, N-back, and SST are presented in Table 2-2. The results for the comparisons of means between the two timepoints (also Table 2-2) indicate that participants performed better at the second testing session compared to the first one on the Bristol ERT and the SST. This means that overall emotion recognition accuracy on the Bristol ERT improved and participants needed less time to successfully inhibit a response on the SST. Hedges' g_{av} Table 2-2 and common language (CL) effect sizes were calculated based on materials provided by Lakens (2013b). The CL effect size for the Bristol ERT indicated that the

likelihood of participants scoring higher in the second session compared to the first was 67%. Similarly, for the SST the likelihood that participants reduced their SSRT in the second session compared to the first is 75%. There was no evidence of a difference in performance on the N-back task, which was also reflected in the CL effect size indicating that the likelihood of and increased *d*'prime in the second session compared to the first was only 52%, where 50% is chance.

Table 2-2

Descriptive statistics for primary outcome measures and paired t-test results

Task (outcome measure)	***n***	Time 1		Time 2		Welch's paired t-test		
		Mean	***SD***	***Mean***	***SD***	***t-value***	***p-value***	g_{av}
Bristol ERT (prop. of total hits)	90	0.693	0.074	0.722	0.074	4.13	< .01	0.389
N-back (*d*'prime)	87	2.858	0.695	2.902	0.707	0.55	.59	0.062
SST ($SSRT_{med}$)	88	255.3	49.03	232.8	37.42	-6.31	< .01	0.511

Note: SD is standard deviation. g_{av} is average Hedges' g corrected for reliability using Pearson's r. ERT is Emotion Recognition Task. Prop. is proportion. SST is Stop Signal Task.

The improvement in scores on the Bristol ERT and SST corresponded to the decrease in ICC2.1 observed for these measurements. To investigate whether the length of time between the first and the second session influenced whether there was an improvement in performance, participants were divided using a median split on length of time between testing sessions. The median time between sessions was 40.5 days, so the sample was split into a short delay between sessions group (27-40 days) and a long delay between sessions group (41-77 days). Comparison of means between the two testing sessions using a paired Welch's t-test showed that in both the short and long conditions participants improved their performance on the Bristol ERT and SST from session one to session two (Table 2-3). Although after correcting for the number of tests conducted using a Bonferroni correction (0.05/9 = 0.005) there was no longer evidence for a difference in scores on the Bristol ERT in the long condition. There was no evidence for a change in performance on the N-back in either condition. The CL effect size estimates for the Bristol ERT indicated that the likelihood of a participant's score being higher in the second session was 71 % in the short delay condition compared to 62% in the long delay condition. See Appendix B (section 8.2) for further details of the effect size calculations.

Table 2-3

Descriptive statistics and paired t-test results split by length of time between sessions

Task (outcome measure)	***n***	**Delay between sessions**	Time 1		Time 2		Paired t-test		
			Mean	***SD***	***Mean***	***SD***	***t-value***	***p-value***	g_{av}
Bristol ERT (prop. of total hits)	45	short	0.697	0.077	0.733	0.075	3.75	<.001	0.466
	45	long	0.690	0.072	0.711	0.073	2.11	0.04	0.285
N-back (*d*'prime)	44	short	2.97	0.677	3.10	0.552	1.16	0.25	0.207
	43	long	2.75	0.704	2.70	0.791	-0.45	0.65	0.066
SST ($SSRT_{med}$)	43	short	253.2	45.4	231.5	35.6	-4.70	<.001	0.522
	45	long	257.5	53.0	234.1	39.6	-4.23	0.001	0.492

Note: SD is standard deviation. g_{av} is average Hedges' g, corrected for reliability using Pearson's r. ERT is Emotion Recognition Task. Prop. is proportion. SST is Stop Signal Task.

A sensitivity analysis was conducted using only cases with full data for all three tasks (see Appendix C, section 8.3). The means and standard deviations were similar and resulted in the same pattern of results when analysed using a paired Welch's t-test.

2.3.5 Standard Error of Measurement and the Smallest Real Difference

The ICC2.1 for each of the cognitive tasks was used to calculate the Standard Error of Measurement (SEM) associated with primary outcome of each task. The confidence intervals around the SEM were used to approximate the Smallest Real Difference (SRD) needed to indicate a true change in an individual's performance between the two testing sessions in ALSPAC (Table 2-4).

Table 2-4

Standard Error of Measurement and Smallest Real Difference of the three cognitive tasks

Task (outcome measure)	**Standard Error of Measurement** (SEM)	**Smallest Real Difference** (SRD)
Bristol ERT (prop. of total hits)	0.05	0.14
N-back (*d*'prime)	0.58	1.6
SST (SSRT)	28.72	79.55

Note: the SEM and SRD are presented in the units of the primary outcome: proportion of total hits for the Bristol *ERT, d'prime for the N-back, and milliseconds for the Stop Signal Task (SST).*

The SEM was used to indicate the error margin around individual scores, which was 5% or approximately 5 trials on the Bristol ERT. The SRD on the Bristol ERT indicated that participants would need to perform differently on 14% of the trials to be confident of a true change in performance. As there were 96 trials in this task participants would have to improve their performance by correctly identifying 13 more emotions or deteriorate in performance by identifying 13 emotions less $(0.14 \times 96 = 13)$.

As the *d*'prime is a standardised measure of discriminability it was harder to translate it into a change on performance on the N-back task. The discriminability score could be increased by getting more hits or reducing the number of false alarms. Given the number of hits and false alarms in this task, the highest *d*'prime value was 3.78 and the lowest was -3.78. Participants would have to shift their discriminability score by 1.6 on this scale to suggest there had been a true change in performance. As a reminder the $SSRT_{med}$, which was the SSRT used in this study, is based on the difference between the median response time on 'go' trials and median stop signal delay time on successful 'stop' trials on the SST. This means the SSRT could be decreased (i.e., improved) through a faster reaction time on the go trials or a longer delay before the stop signal is given on successful stop trials. The standard error of measurement associated with and individuals SSRT was approximately 29ms. The SDR indicated that participants would need to decrease or increase their SSRT by approximately 80ms to suggest a true change in performance.

2.3.6 Individual emotions on the Bristol ERT (secondary analysis)

For the Bristol ERT the reliability coefficients for each emotion were calculated using an unbiased hit rate (H_u). The reliability coefficients were low across the board, indicating poor/low reliability of the by emotion outcome measure in this sample (Table 2-5). The lowest reliability in performance was for sad faces, where the 95 % confidence interval did not extend beyond the poor range. The confidence intervals for all the other emotions suggested poor to moderate reliability, although all of them except disgust were primarily in the poor range. Again, the ICC3.1 and Pearson's r were the same for each emotion. The ICC2.1 was marginally lower for happy, fearful, surprised, and disgusted face recognition accuracy indicating that there may have been a systematic shift in accuracy between sessions for these emotions, but not for sad or angry face recognition accuracy.

Table 2-5

Reliability coefficients for each individual emotion form the Bristol ERT

Bristol ERT by Emotion (outcome = H_u)	Intraclass Correlation Coefficient (ICC)					Pearson's r		
	df	**ICC2.1**	**95% CI**	**ICC3.1**	**95% CI**	**df**	**r**	**95% CI**
Happy	89	0.35	0.16, 0.52	0.37	0.17, 0.53	88	0.37	0.17, 0.53
Sad	89	0.24	0.04, 0.43	0.24	0.04, 0.43	88	0.24	0.04, 0.43
Angry	89	0.44	0.26, 0.59	0.44	0.26, 0.59	88	0.44	0.26, 0.60
Fear	89	0.43	0.23, 0.59	0.46	0.28, 0.61	88	0.46	0.28, 0.61
Surprise	89	0.36	0.17, 0.52	0.37	0.18, 0.54	88	0.37	0.18, 0.54
Disgust	89	0.50	0.33, 0.64	0.51	0.34, 0.65	88	0.51	0.34, 0.65

Note: H_u is the unbiased hit rate as defined by Wagner (1993). df is degrees of freedom. CI refers to 95% Confidence Interval.

Descriptive statistics and results for a paired t-test comparing mean accuracy for each emotion at the two timepoints are presented in Table 2-6. The mean accuracy scores showed a trend towards improvement in emotion recognition accuracy. The results for the paired Welch's t-test indicated that only happy, fearful, and surprised faces were recognised more accurately at the second session compared to the first. Although, when a Bonferroni correction was applied to account for the number of tests conducted, the difference in accuracy score for surprised faces no longer met the significance threshold of 0.008 (0.05/6 = 0.008). As these results were based on an unbiased hit rate (H_u) the increase in accuracy for happy and fearful faces could be explained by an increase in correct identification of happy and fearful faces, or a decrease in the number of incorrect identifications of these emotions. These results suggested that the changes in overall emotion recognition may have been driven by changes in recognition accuracy for specifically happy and fearful faces, although this was not formally assessed.

Table 2-6

Summary of outcome measures and paired t-test for each emotion from the Bristol ERT

Bristol ERT by Emotion (outcome = H_U)	Time 1		Time 2		Paired t-test		
	Mean	***SD***	***Mean***	***SD***	***df***	***t-value***	***p-value***
Happy	0.560	0.106	0.595	0.109	89	2.77	< 0.01
Sad	0.562	0.116	0.583	0.120	89	1.39	0.17
Angry	0.570	0.168	0.598	0.148	89	1.55	0.12
Fear	0.319	0.212	0.405	0.220	89	3.65	< 0.01
Surprise	0.449	0.116	0.485	0.115	89	2.57	0.01
Disgust	0.609	0.145	0.631	0.141	89	1.48	0.14

Note: HU is unbiased hit rate as defined by Wagner (1993). df is degrees of freedom. SD is standard deviation.

2.4 Discussion

The main aim of this study was to evaluate the test-retest reliability of overall emotion recognition accuracy (total hits) on the Bristol ERT completed by the ALSPAC cohort. The reliability coefficients for the Bristol ERT indicated that reliability in this sample was moderate/marginal. The reliability coefficients for the visuoverbal 2-back version of the N-back task were suggestive of poor reliability and measurements made using Stop Signal Task seemed to have good test-retest reliability. The results for the Bristol ERT were evaluated in the context of the current literature and the wider implications for reliability of measurement in ALSPAC discussed in considering the reliability of all three cognitive tasks in this sample. See Appendix A (section 8.1) for a discussion of the reliability of the N-back and Stop Signal Task.

2.4.1 Test-retest reliability estimates for the Bristol ERT

To the best of our knowledge, this was the first study investigating test-retest reliability of overall emotion recognition accuracy on the Bristol ERT. The reliability estimates for the Bristol ERT in this sample were in the threshold between low and marginal according to the thresholds outlined by Strauss and colleagues (2006). Arguably 0.7 is the minimum acceptable test-retest reliability threshold for tasks used to assesses psychological outcomes (Strauss et al., 2006). The Bristol ERT clearly does not meet this standard in the ALSPAC cohort. Confidence intervals around the ICC reliability estimates suggested that test-retest reliability in this sample could be poor and was moderate at best according to the thresholds proposed by Koo and Li (2016). These results indicated that the Bristol ERT is not well suited for tracking changes in individual performance as it would be difficult to differentiate true change from variance in measurement. That does not mean that the Bristol ERT is unable to detect group differences as outlined by Hedge and colleagues (2018).

The measure of agreement between scores (ICC2.1) on the Bristol ERT being slightly lower than the measures of consistency (ICC3.1 and Pearson's r) could be explained by the systematic improvement in scores in the second testing session compared to the first. The fact that the improvement in performance seems to be driven by participants who had a shorter delay between sessions was suggestive of a practice effect. This means that measures of consistency (ICC3.1 and Pearson's r) may be more appropriate when evaluating test-retest reliability on the Bristol ERT when the time between testing sessions is short e.g., within one to two months (Parsons et al., 2019).

The reliability estimates for consistency were both in the moderate/marginal range, so research studies using the Bristol ERT to assess change in emotion recognition over time should consider the impact low reliability could be having and evaluate whether the Bristol ERT is suitable for the intended use. That being said, more research is needed to establish pooled estimates across different samples and evaluate reliability of the Bristol ERT in other populations.

The effect size estimates for the difference in scores between sessions were corrected for reliability using Pearson's r and indicate that the likelihood of scores changing between sessions was only slightly higher than chance. Despite there being a statistically reliable difference in performance between sessions the size of this effect was small. In the context of this study, a small effect size was desirable, because a true change in performance across testing sessions was not anticipated, although it is worth noting that the group effect size was small but individual variation in performance could have been big. Longitudinal research effect sizes are often smaller than in cross-sectional research as baseline performance is controlled for (Adachi & Willoughby, 2015) and more time between measurements can lead to more variation. It would be worth investigating the effect size needed for a meaningful change in performance i.e., practical significance (Lakens, 2013b). The Standard Error of Measurement (SEM) suggested there was substantial variation in individual scores. In turn, the Smallest Real Difference (SRD) shows that a large shift in performance is required to be confident that there is a true change in performance (Vaz et al., 2013). However, the SEM and SRD may be inflated due to the improvement in performance observed between the first and second testing sessions on the Bristol ERT.

The ICC2.1 in this study was lower than the one reported by Bland and colleagues (2016) for the Bristol ERT. However, it is not possible to draw a direct comparison as the outcome measure used in their study was different to the one used here. Equally, the reliability estimate of r = 0.84 reported for the emotion recognition task based on Ekman and Friesen face stimulus set used by Palmer and colleagues (2018) is not comparable. One, the outcome measure they used was only based on negative emotions, and two, all stimuli were presented at 100% intensity. The stimuli being presented at varying levels of intensity on the Bristol ERT are likely to have increased variation in individual performance and thus could have contributed to a decrease in the reliability estimates. By emotion test-retest reliability was assessed using an unbiased hit rate in the current study. The reliability estimates for individual

emotions were low across the board in the current study, which is in line with the reliability estimates based on unbiased hit rate reported by Cecilione and colleagues (2017). These results suggest that overall emotion recognition accuracy is a more reliable measure of emotion recognition than individual emotion recognition accuracy on the Bristol ERT. Notably, the reliability estimates for individual emotion recognition accuracy were much better when analysed using latent growth modelling (Cecilione et al., 2017) or when based on a discriminability index (Adams et al., 2016; Grier, 1971). These may be more appropriate outcome measures when evaluating change in performance over time for individual emotions but were not included here as trial level data was not available for this study.

It is important to note that an underlying assumption in the assessment of test-retest reliability is that the true score has stayed the same. A long test-retest interval might result in a true change, which would increase variation in performance and thus impact reliability estimates. Whilst there is no recommended time between test intervals studies commonly report an interval of two to fourteen days (Streiner et al., 2008). The studies investigating test-retest reliability of emotion recognition discussed here all fall within that window, except for the study by Cecilione and colleagues (2017) where the interval was two to five weeks. In ALSPAC the test-retest interval is at least 27 days, so much longer, consequently it is possible that true change in performance could be decreasing the reliability estimates. A long testing interval is of particular concern if state more so than trait attributes are thought to influence performance on a test (Polit, 2014). There is evidence that performance on the Bristol ERT is influenced by both state and trait components of anxiety (Attwood et al., 2017). It would be beneficial to investigate test-retest reliability using a shorter testing interval to investigate this further, for example by comparing performance a couple hours later to a week later.

2.4.2 Strengths and Limitations

Given that reliability is a property of measurements made within a given sample, the reliability estimates calculated using ALSPAC data are likely to generalise to the wider cohort. Although, the reliability estimates could be partially biased given that participants who return for a second testing session were a self-selecting sample and invited based on proximity to Bristol and willingness to engage (Polit, 2014). It is possible that homogeneity of the sample (e.g. narrow age range and willingness or ability to return for a second testing session) could reduce reliability estimates (Hedge et al.,

2018). Further, the SEM for each of the tasks indicated there was variability around individual scores (Parsons et al., 2019). Low reliability estimates in this sample could be due to participants completing the cognitive tasks at the end of a long day of testing. Lack of attention or fatigue could be affecting the quality of the data collected and increase variability due to random error. This is especially true for the first testing session, which included a larger variety of tests, not all of which were repeated at the second testing session. This could also partially explain the improvement in performance observed between the first and the second testing session on the Bristol ERT and SST.

Low reliability for group differences can be mitigated through large sample sizes, because studies are likely to have sufficient power after taking reliability into consideration (Parsons et al., 2019). At the first testing session as part of the age 24 clinic in ALSPAC, 4026 participants completed the cognitive tasks evaluated in this study and of these more than 85 had corresponding data at the second session. A strength of using the data collected as part of ALPAC for this study, is that the sample size was relatively large compared to many of the studies reporting test-retest reliability for the three cognitive tasks included in this study. Further, a lot of the research conducted using data for these tasks collected at the first testing session is likely to have sufficient power despite the low test-retest reliability. Nevertheless, it would be prudent to consider the impact reliability might be having in this cohort as opposed to assuming that it is not an issue. The fact that reliability estimates for the N-back and SST were within the expected ranges based on the current literature (Appendix A, section 8.1) there was no evidence that quality of data in ALSPAC was the reason for low test-retest reliability estimates.

2.4.3 Implications and future directions

Regardless of what was causing the high variance in performance and in turn low reliability coefficients for the three cognitive tasks, it is important to consider the implications for future research in ALSPAC. Research investigating change over time using the SST or N-back, which have been administered at several timepoints in ALSPAC, should begin accounting for low test-retest reliability. Equally, studies controlling for individual differences should evaluate the utility of including the N-back, SST, or Bristol ERT or take into consideration the potential impact of low reliability when drawing inferences based on that data. Caution should be applied in using these reliability estimates presented in this study for other samples or populations, although it would be

advisable to consider the impact low reliability could be having and calculate reliability in that sample or population where possible.

The results from this study highlight the effect that use of different outcome variables could be having on reliability estimates. Tasks in a clinical setting are required to have clearly defined outcome measures that are calculated in the same way every time (Kessels, 2019). Tasks used in a research setting would benefit from a better understanding of the outcome measures used to allow for better comparisons across studies. Consistency in outcome measures would also allow for pooled estimates of test-retest reliability to be established as proposed by Parsons and colleagues (2019). In terms of the Bristol ERT the results of this study indicate that the unbiased hit rate used to assess individual emotion recognition accuracy may not be a suitable measure to assess change in performance over time. Similarly, overall emotion recognition accuracy based on the total number of hits was not a reliable measure of individual change in performance over time, with a true difference in score being associated with a difference of 13 hits on the Bristol ERT. Further, cognitive tasks are increasingly being used as outcome measures in research or cognitive trials, without considering the size of the effect required to indicate a meaningful change in performance. Equivalence testing has been proposed as a formal way in which to assess whether there is a meaningful change in performance (Lakens, 2017). Future research should consider what effect size would constitute a meaningful change performance factoring in reliability, given that the Bristol ERT was associated with high variance in measurement.

Chapter 3 Construct Validation for the Bristol Emotion Recognition Task

3.1 Introduction

The previous chapter introduced the topic of psychometrics and their importance in both research and clinical settings. It went on to address the concept of reliability because, as stated previously, reliability is necessary but not sufficient to infer validity (Cook & Beckman, 2006). The premise behind this statement is that you cannot claim to effectively measure something if the task being used is unreliable, but equally a task could be reliable but not measure what is intended to be measured. This means that the concept of validity addresses the extent to which a task (or test, scale) measures what it intends to measure (Streiner et al., 2015). This chapter will further explore the concept of validity and how validity can be established. Subsequently a validation study using the Bristol Emotion Recognition Task (ERT) will be presented. The aim was to evaluate whether valid inferences about emotion recognition can be made based on overall emotion recognition accuracy (total hits) measured using the Bristol ERT. Performance on the Bristol ERT was compared to two well validated measures of emotion recognition; an Emotion Recognition Task using dynamic stimuli (Dynamic ERT) developed by Montagne et al. (2007), and the Emotion Evaluation Test from The Awareness of Social Inference Test (TASIT; McDonald et al., 2003). The study presented in this chapter was registered on the Open Science Framework and sections are taken from the pre-registered protocol (10.17605/OSF.IO/JCFD6).

3.1.1 What is validity?

Validity in the context of psychological testing is defined as "the degree to which evidence and theory support the interpretation of test scores for proposed [entailed] uses of tests" in 'Standards for educational and psychological testing' (APA et al., 1999; APA et al., 2014). This definition represents a consensus based on our current understanding of the concept agreed upon by a joint committee from the American Educational Research Association, American Psychological Association, and National Council on Measurement. It has been periodically updated by them since they first published a definition in 1954 (Newton & Shaw, 2013). It is worth noting that the above definition has been described as ambiguous by experts in the field of validity and there is still ongoing debate about the concept of validity (Camargo et al., 2018). A Delphi study looking for consensus on the concept of validity found that the definition of

validity with highest consensus among participating experts was “the degree to which collected evidence, theory, and logical argument support the intended inferences to be made from test scores” (Camargo et al., 2018). Both definitions make it clear that validity is not considered to be an inherent property of a task or test but relates to interpretations made using the scores obtained using that task/test. The inferences drawn based on these scores are what can be considered valid or invalid (Streiner et al., 2015). Newton and Shaw (2013) found that validity is still commonly used to describe tests, measures, or items in the validity literature. This could be due to an implicit understanding that a ‘test/item is valid’ only in the context that it was used in and written about. For the purpose of this thesis it is made explicit that validity is a property of inferences made based on test scores and not the test used. The validity of inferences can vary based on the population being tested, context a test is used in, or intended purpose (Streiner et al., 2015). For example, inferences made based on scores collected in a healthy population may not be valid in a TBI population. Validation studies would be required for both these contexts. This means that test validation is an ongoing process and not a dichotomous decision (Bannigan & Watson, 2009; Streiner et al., 2015).

Historically, validity has been divided into different types, for example, criterion validity, content validity, and construct validity (Bannigan & Watson, 2009; Cronbach & Meehl, 1955; Streiner et al., 2015). There was a shift in thinking in 1980 when Messick argued that there is only one type of validity, namely construct validity. Validity is now viewed as a unitary concept and using descriptors to talk about different types of validity can be misleading (Messick, 1980; Newton & Shaw, 2013; Streiner et al., 2015). In other words, an inference can be either valid or invalid and should not be described as a particular type of valid. Despite this change in thinking about validity, descriptors about types of validity can still be commonly found in the literature (Newton & Shaw, 2013). A possible explanation is that whilst there are not different types of validity, there are lots of different ways and reasons to assess validity. These different validation approaches are likely to have led to the notion that there are different types of validity (Streiner et al., 2015). When Cronbach and Meehl (1955) introduced the concept of construct validity, they in fact talk about four different types of validation but then refer to them as predictive, concurrent, content, and construct validity. It is possible that when referring to different types of validity in the literature the implicit understanding is that the descriptors are a reference to the process of validation and not the concept of validity (Newton & Shaw, 2013).

In sum, validity is a property of the inferences based on test scores and consequently assessing the validity of tests is on ongoing process. Whether inferences made are valid is dependent on context and intended purpose of the test and should not be generalised or assumed to transfer to different contexts. For example, valid inferences about emotion recognition based on the Bristol ERT could be possible in a neurologically healthy population but not a TBI population. Finally, validity is a unitary concept and not composed of different types of validity. When referring to 'types' of validity in this thesis it is to be understood that this references the validation approach used to evaluate validity of inferences made.

3.1.2 Validation approaches

Validity is not an attribute that can be measured but has to be judged given the evidence available in the context of current theory (Cook & Beckman, 2006; Messick, 1980). The type of evidence required to make that judgement will change based on the inferences that will be made and current understanding of theory. Simply, validation is an ongoing process of hypothesis testing (Streiner et al., 2015). If responses on tests are associated with other factors as predicted, this provides evidence that inferences based on these test scores are valid (Cronbach & Meehl, 1955; Streiner et al., 2015). The hypotheses tested should answer a question relevant to the intended purpose of the test and the validation approach used will vary accordingly.

Three commonly used categories for validation approaches are criterion based validation, content validation and construct validation (Streiner et al., 2015). Criterion based validation can be used when an objective criterion is available (Cronbach & Meehl, 1955; Streiner et al., 2015). It is used to check whether a test adequately predicts or is associated with a criterion of interest (often referred to as predictive or concurrent validity). For example, one could check whether performance on the Bristol ERT predicts whether someone is in a relationship after having TBI. The purpose of content validation is to evaluate whether a measure has included only relevant and sufficient items required to assess a given construct (Bannigan & Watson, 2009; Streiner et al., 2015). For example, by investigating whether items load on the same factor structure or assessing variability in responses on a given item. Arguably, face validation is a facet of content validation not assessed with statistical measures but by considering about whether the test seems relevant and appropriate for its intended use (Bannigan & Watson, 2009). Construct validation was first introduced by Cronbach and Meehl in

1955. They argued that in psychological science construct validation is often the most appropriate approach because when assessing cognition or behaviour an objective measure of a variable is often not available (Cronbach & Meehl, 1955; Stone, 2019). For example, emotion recognition is a skill that we assume people to possess but it is not directly measurable. The construct is hypothetical and should be rooted in theory, so validation is based on linking evidence back to theory using a priori predictions (Cook & Beckman, 2006). This means that construct validation can be approached in many different ways and unlike criterion based validation or content validation it cannot be evaluated with just a couple of studies (Streiner et al., 2015). The hypotheses tested in any given study will depend on the question of interest and evidence of validity is obtained if results support the hypotheses made (Streiner et al., 2015).

3.1.3 Emotion Recognition Tasks

The construct of interest in this study is emotion recognition, which is assumed to be an inherent property that people have and can be measured using emotion recognition tasks (Cronbach & Meehl, 1955). A range of different tasks have been used to measure emotion recognition (Paiva-Silva et al., 2016) but as was discussed in the introductory chapter (1.3.1 Measuring Emotion Recognition) many tasks have insufficient data on validity and lack normative data (Howieson, 2019). The Awareness of Social Inference Test (TASIT, McDonald et al., 2003) is a well validated test battery to assess social cognition and has been recommend as a measure in a TBI population (Honan et al., 2016; Wallis et al., 2021). The Emotion Evaluation Test (EET) is the subscale of TASIT used to measure emotion recognition. The stimuli in this task are a set of short video clips, which arguably improves the validity of inferences made regarding real life effects (McDonald et al., 2004) as the stimuli are dynamic and participants can use body language and movement in addition to facial expressions to provide information about emotion (Atkinson et al., 2004). However, this approach makes it difficult to manipulate intensity of emotions presented, which could result in ceiling effects that mean more subtle difficulties or biases in emotion recognition are missed. A potential solution is to use a morph sequence of static images to make short video clips as used in the Emotion Recognition Task developed by Montagne and colleagues (2007). The Dynamic ERT, as it will be referred to in this thesis, has normative data available (Kessels et al., 2014) and has been used to investigate emotion recognition in a TBI population (Rosenberg et al., 2015; Rosenberg et al., 2014). Performance on these to tasks will be compared to performance on the Bristol ERT in this study.

3.1.4 Current Study: Aim

The aim of this study is to provide evidence supporting the validity of inferences about emotion recognition made using scores on the six-emotion version of the Bristol ERT. Whether the Bristol ERT adequately measures the construct of interest will be evaluated using a series of correlations. Performance on the Bristol ERT will be compared to emotion recognition performance on the EET from a short version of TASIT (TASIT-S) and the Dynamic ERT to show that the Bristol ERT is equivalent to these tests despite using static stimuli. The Bristol ERT should have a strong correlation with these tasks (convergent evidence), but weak or no correlation with tasks that are trying to measure different constructs (divergent evidence). The study design is based on a methodology for construct validation that was first introduced by Campbell and Fiske (1959). They proposed that a multitrait-multimethod matrix should be used to assess whether a task is effectively measuring the construct it is trying to measure. Tasks trying to measure the same construct using varying methodologies should show a strong correlation (often referred to as convergent validity). Similarly, tasks trying to measure a different construct by the same or different methodologies should show a weak correlation or no correlation (often referred to as divergent validity). Evidence that a task of interest effectively measures the construct it is trying to measure is obtained if the pattern of correlations within the multitrait-multimethod matrix matches predictions made based on prior evidence and theory. This in turn would provide evidence that the inferences made based on performance on that task are valid. Correlation matrices combining convergent and divergent evidence are a commonly used validation approach (Streiner et al., 2015) and are useful in the development of new tasks such as the Bristol ERT.

3.1.4.1 Convergent evidence

A high correlation of the Bristol ERT with other emotion recognition tasks would provide evidence that they are tapping into the underlying construct of emotion recognition. This would provide support for the validity of inferences made about emotion recognition based on performance on the Bristol ERT. The three tasks compared were the Bristol ERT, the Dynamic ERT (Montagne et al., 2007) and EET from TASIT-S (Honan et al., 2016). The Dynamic ERT and EET from TASIT-S were chosen because they aim to measure the same construct as the Bristol ERT but differ in respect to the types of stimuli used. Further, there is evidence for the validity of inferences

made about emotion recognition using these tasks and both have normative data available (Honan et al., 2016; Kessels et al., 2014; McDonald et al., 2006; McDonald et al., 2004; McDonald et al., 2018; Montagne et al., 2007; Rosenberg et al., 2015). Normative data for the Dynamic ERT is available from 373 healthy participants across a series of studies (Kessels et al., 2014). Overall performance on TASIT-S is strongly correlated with TASIT, and the correlation of the EET in the two versions is r = .87. Normative data for TASIT-S is available from 649 healthy participants recruited in Australia and the USA (McDonald et al., 2018).

Alexithymia is a disorder where people are unable to effectively identify and express their emotions (Lesser, 1981; Taylor & Bagby, 2004). Evidence suggests that performance on emotion recognition tasks is linked to or even dependent on the ability to express emotion (Parker et al., 1993; Wearne et al., 2019). To investigate this the twenty item Toronto Alexithymia Scale developed in 1992 (Bagby, Parker, et al., 1994; Bagby, Taylor, et al., 1994) was included in this study. If ability to express emotions is crucial for effective emotion recognition the emotion recognition tasks should have a strong negative correlation with this measure of alexithymia.

3.1.4.2 Divergent evidence

Performance on any cognitive test is likely to capture some aspect of general cognitive ability such as working memory, but general cognitive ability should not fully explain test performance. Digit Span tasks are included in the Wechsler Adult Intelligence Scale as a measure of working memory (Bowden et al., 2013; Hilbert et al., 2015; Ramsay & Reynolds, 1995). In this study, a visual version of the Digit Span Backwards was used as a general measure of working memory. There has been some debate about what aspect of working memory is captured by the Digit Span Backwards (Hilbert et al., 2015) or whether it is better described as a measure of short-term memory (St Clair-Thompson, 2010). Regardless of what aspect of general cognition is captured by the Digit Span Backwards task, it should differentiate from performance on the emotion recognition tasks. This means a low correlation of the Bristol ERT with this task would suggest that variation in performance on the Bristol ERT is not solely due to differences in general cognitive ability.

The emotion recognition tasks in this study require identification of emotions based on facial expressions, which means that performance is likely to be influenced by individual differences in face perception. The EET from TASIT has been shown to

correlate with the Benton Face Recognition Test at r = 0.45 (McDonald et al., 2006). To evaluate whether performance on the Bristol ERT is capturing emotion recognition and not purely face perception the Glasgow Face Matching Test (GFMT) was included as a measure of face discrimination in this study (Burton et al., 2010). A positive correlation in performance between the Bristol ERT and GFMT is to be expected. If the Bristol ERT shows a higher correlation with the other emotion recognition tasks than with the GFMT this would support the validity of inferences about emotion recognition made based on Bristol ERT scores.

3.1.5 Current Study: Hypotheses

The EET from TASIT-S was shown to have a correlation of r = 0.69 with an emotion recognition task based on the Ekman and Friesen pictures (Friesen & Ekman, 1976; McDonald et al., 2006). The correlation of the Bristol ERT with itself at a later timepoint was r = 0.6, so this estimate was used for the power calculation and as a predicted correlation coefficient for performance on the three emotion recognition tasks. Given that unlike the EET, the Bristol ERT and Dynamic ERT both use face stimuli they are likely to be more highly correlated with each other. Consequently, the two primary hypotheses for this study were:

H1: The three emotion recognition tasks will have a positive correlation coefficient of around 0.6 when correlated with each other.
H2: The Bristol ERT and the Dynamic ERT will be more highly correlated with each other than with the EET from TASIT-S.

The correlation between the three emotion recognition tasks should be higher with each other than with the measures of working memory and face discrimination. McDonald and colleagues (2006) reported a correlation of r = 0.25 between the EET from TASIT and the Digit Span measure (based on both the digit span forwards and backwards) from the third version of the Wechsler Adult Intelligence Scale. In the same study the correlation between the EET from TASIT and a face recognition task was r = 0.45 (McDonald et al., 2006). Given these correlation coefficients and the expected interaction between the measures of emotion recognition, working memory, and face discrimination based on current theory, the secondary hypotheses were:

H3: The pattern of correlation of the emotion recognition tasks will show the lowest correlation with the Digit Span Backwards, a slightly greater correlation with the Glasgow Face Matching Test, and the highest correlation with each other.

H4: The correlation coefficient of the emotion recognition tasks with the Digit Span Backwards will be less than 0.4.

Finally, it is possible that the performance on the emotion recognition tasks is dependent on ability to express emotions. This means that performance on all three emotion recognition tasks would be negatively correlated with scores on the Toronto Alexithymia Scale because a higher score on that scale indicates decreased ability in identifying and expressing emotions. The final hypothesis for this study was:

H5: Performance on the emotion recognition tasks will be negatively correlated with the Toronto Alexithymia Scale.

3.2 Method

3.2.1 Design

This is an observational cross-sectional study investigating the construct validity of the Bristol ERT. Participants were recruited online and asked to complete three emotion recognition tasks, a working memory task, a face discrimination task and a self-report questionnaire about ability to process emotions (alexithymia). The order of presentation of the three emotion recognition tasks was randomised. The hypotheses for this study and proposed analyses were preregistered on the Open Science Framework (10.17605/OSF.IO/JCFD6). Ethics approval was obtained from the School of Psychological Science Research Ethics Committee at the University of Bristol (103142).

3.2.2 Participants

Participants were recruited on Prolific (https://www.prolific.co/participants). They were able to access the study information and sign-up if they matched screening criteria set on Prolific. The criteria were set to identify participants 18 years old and above with no history of neurological conditions. See Appendix A (section 8.4) for screening questions. Participants were also asked to confirm they matched the following inclusion and exclusion criteria as part of the consent form before starting the study.

Inclusion criteria

- Above the age of 18
- Normal or corrected to normal vision
- Access to a desktop computer or laptop to for the study

Exclusion criteria

- Currently under the influence of substances that can substantially alter perception (e.g. psychoactive drugs or excessive alcohol)
- History of psychiatric or neurological conditions

A sample size calculation was conducted to establish the number of participants required to address the study hypotheses. The principal consideration was to establish correlation between two tasks with sufficient precision to confidently distinguish between correlations that were expected to be convergent versus divergent. A priori sample size calculation was conducted in GPower 3.1 (Faul et al., 2007) using precision estimates for the expected correlation between the emotion recognition tasks. For the expected Pearson correlation of around 0.6, 182 participants were required to establish that the correlation is within 0.2 of that estimate at 95% power and an error probability of 0.05. Based on this calculation the aim was to recruit 185 participants.

3.2.3 Measures

The tasks and questionnaires were coded in and presented through Gorilla Experiment Builder (Anwyl-Irvine et al., 2019).

3.2.3.1 Demographics questionnaire

Participants were asked to provide information about sex, age, years of education and given the option to declare reasons why their emotion recognition might be affected. The last question stated "This is an attention check. To show you have read this question select Yes.". Participants who selected 'No' were considered to have failed the attention check Additionally, demographic information available on Prolific was downloaded for the participants that completed the study to cross reference responses if missing and check participants' nationality.

3.2.3.2 Bristol Emotion Recognition Task (Bristol ERT) – short version

A total of 96 static facial expressions are presented in a randomised order. Participants are asked to label the stimuli using one of the six 'basic' emotions, angry, sad, happy, fearful, disgust, or surprise. Each emotion is presented at eight levels of intensity using both male and female stimuli (For a detailed description see section 1.3.2 in the introduction). A 'total hit rate' was calculated using the total number of correct identifications across all emotions as a measure of overall emotion recognition accuracy (total hits). Normative data for this task is not currently published.

3.2.3.3 Dynamic Emotion Recognition Task (Dynamic ERT)

The Dynamic ERT was developed by Montagne and colleagues (2007) and is available at: http://www.metrisquare.net/metrisquare/emotion-recognition-test/. Like the Bristol ERT, it is a six-alternative forced-choice emotion recognition task used to investigate identification of emotions from facial expressions. However, the stimuli in this task are comprised of dynamic morph sequences presented as video clips, instead of static facial stimuli, hence it is referred to as Dynamic ERT in this study. The shortest sequence is comprised of images from neutral to 40% intensity, followed by sequences from neutral to 60%, neutral to 80%, and neutral to 100% intensity and the video clips range from around 1 to 2 seconds in length. The six emotions included are happiness, sadness, anger, fear, surprise, and disgust. There are four sets of images, (2 male and 2 female), so 96 stimuli are presented in total. Participants completed three practice trials

from a separate stimulus set and then the test stimuli are presented in a set order of four blocks from lowest to highest intensity. For each trial, participants viewed a video clip and then gave a response whilst being able to view the final image of the clip. The primary outcome for this study was total hits (i.e. correct identifications) as a measure of overall emotion recognition and secondary outcomes were hits per emotion. Normative data is available for healthy controls recruited in the Netherlands, Ireland, Germany and Australia (Kessels et al., 2014).

3.2.3.4 The Awareness of Social Inference Test – Short (TASIT-S): Part 1

Part 1 of TASIT-S is an Emotion Evaluation Test (Honan et al., 2016) used to assess overall emotion recognition accuracy using 10 video clips. Duration of the video clips (aka vignettes) ranges from 15 to 60 seconds and they show one or two actors engaged in dialog or an activity (if two actors are in the video, the participants are given instructions regarding which person to focus on). The content of the scripts is neutral, but the facial expression and body language are used to present one of six emotions: happy, sad, angry, fear, disgust or neutral. After watching the full clip, participants are asked to choose which emotion was presented from a list of seven emotions: happy, sad, angry, fear, disgust, neutral and surprise. Surprise is an option because it is included in the original version of TASIT, even though none of the video clips in TASIT-S show surprise. Participants are given a practice video clip with feedback followed by the 10 test video clips. The stimuli are presented in a set order, in accordance with TASIT-S manual. The Emotion Evaluation Test from TASIT-S can only be used to assess overall emotion recognition, not emotion specific performance because the task was developed as a screening tool and most emotions are only presented once. Total number of correct identifications was used as the primary outcome for this task. Normative data is available for Australian and US populations (McDonald et al., 2018).

3.2.3.5 Digit Span Backwards (DSB)

The Digit Span Backwards task is a simple measure of attention and working memory. A visual version of the task was used from the open materials available on Gorilla (Massonnie, 2019, https://gorilla.sc/openmaterials/36699) and adapted for the purposes of this study. Participants are presented with digits between 0 and 9 in a randomised order at a rate of one per second. They are asked to recall the digits in reverse order when prompted by entering the numbers using a keypad presented on the

screen. Two practice trials are presented using a digit span of two and participants were given feedback based on their response. The task trials start with a digit span of three and a digit is added every three trials. The task ends when participants respond incorrectly to two out of three trials of a given digit span or participants have completed the trials with a digit span of 10. The primary outcome for this study is length of the longest digit span correctly recalled in reverse order.

3.2.3.6 Glasgow Face Matching Test (GFMT) – short version

The GFMT assesses people's ability to recognise unfamiliar faces. In the short version 40 pairs of faces are presented and participants are asked to make a judgment as to whether the faces presented are of the same person or two different people. There are 20 trials that show faces for the same person and 20 trials showing faces of different people. The order in which the faces are presented is randomised and the faces remain visible on the screen until participants have made a judgment. Number of hits across all trials is used as a measure of overall accuracy. UK normative data is available (Burton et al., 2010).

3.2.3.7 Toronto Alexithymia Scale (TAS)

The 20 item TAS was used in this study, which can be separated into three subscales (Bagby, Parker, et al., 1994; Bagby, Taylor, et al., 1994). These include: 1) difficulty in identifying feelings, 7 items, 2) difficulty in describing feelings, 5 items, and 3) externally orientated thinking, 8 items. The aim is to capture both affective processing and cognitive strategies that contribute to the construct of alexithymia (Bagby & Taylor, 1997). Items are rated using a 5-point Likert scale whereby 1 = strongly disagree and 5 = strongly agree. This means the minimum score is 20 and the maximum score is 100, but 5 of the items are negatively keyed and are reverse scored for the total score. The TAS-20 was developed to be used as a continuous variable, but cut-off scores are available. A score of equal to or less than 51 is classed as non-alexithymia, and equal to or greater than 61 is classed as alexithymia and scores of 52 to 60 are possible alexithymia (Bagby & Taylor, 1997). Normative data from a Canadian population is available from Parker et al. (2003). In addition to the 20 items an attention check was added halfway through this questionnaire. The attention check included the following instruction "This is an attention check. Please respond with four - Agree." If participants did no select option four they were considered to have failed the attention check.

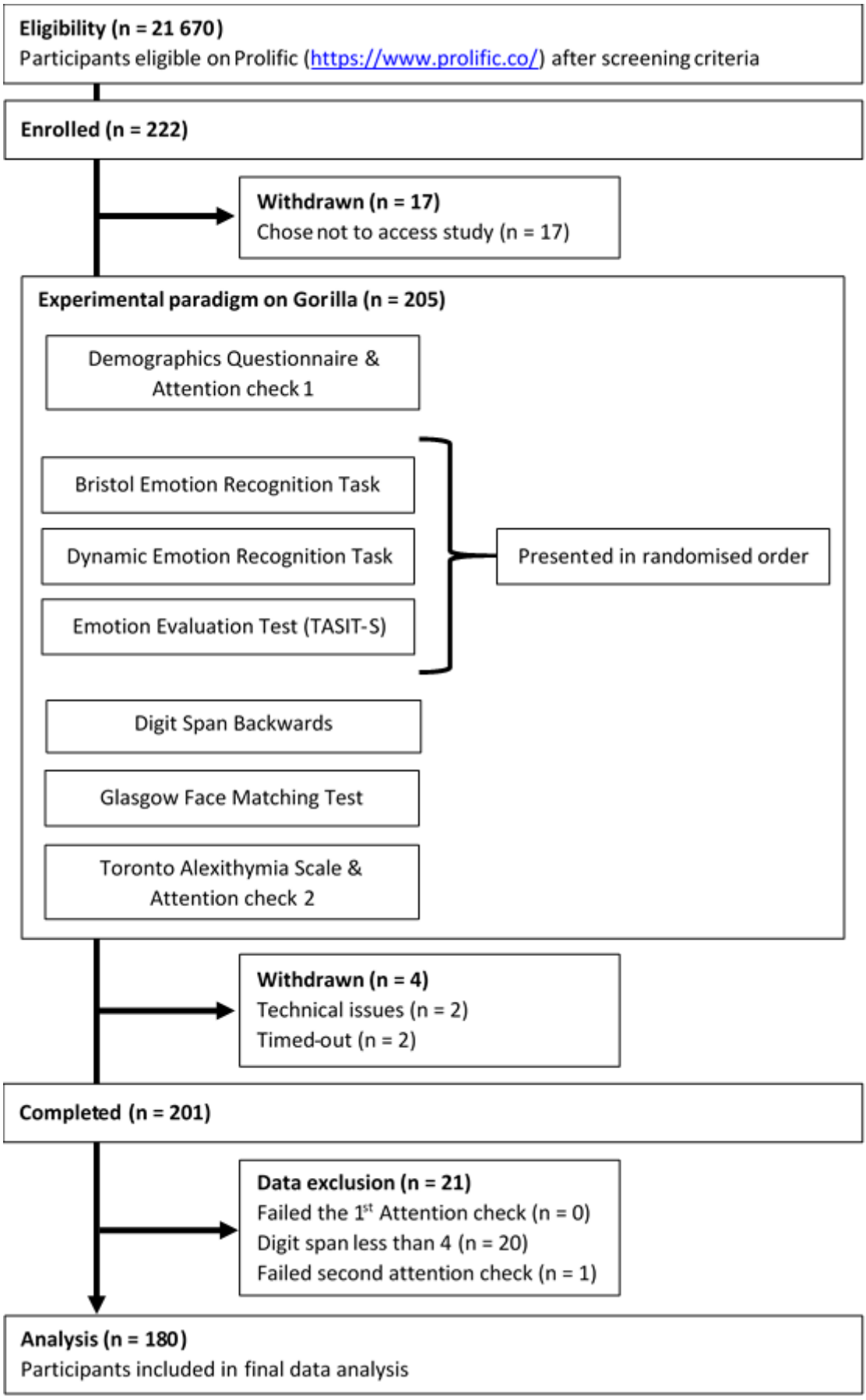
3.2.4 Procedure

Participants were given information about the study on Prolific and had opportunity to contact the research team before signing up to take part. Participants who chose to enrol in the study were directed to Gorilla using a link and asked to review the information sheet before completing an electronic consent form. Once enrolled they had a maximum of 115 minutes to complete all the tasks before being automatically rejected from the study. All participants were asked to complete every task, but the order in which the three emotion recognition tasks were presented was randomised across participants. See the Consort diagram (Figure 3-1) for an overview of the study procedure. Before starting the tasks and questionnaires participants were asked to complete the study in one session and were given the option of taking short breaks between the tasks. They were also instructed to check their audio settings to ensure that they had sound required to play the video clips for the EET from TASIT-S. Task specific instructions were given at the start of each task. At the end of the study, participants were asked to confirm that the data could be used for the purpose of this study and be made available as open data. They were then given a completion URL directing them back to Prolific and reimbursement for the study was authorised by the research team. All participants who completed the study were reimbursed £5.50 for their time regardless of whether the data was used in the analysis.

3.2.5 Data Collection and Screening

Gorilla Experiment Builder was used to manage data collection and storage (Anwyl-Irvine et al., 2019). Recruitment happened in four stages. Initially only 3 participants were recruited, and the data was checked for completeness. A further 17 participants were recruited after which a data quality check was conducted. No issues were identified so a further 165 participants were recruited to a total of 185 participants set based on a sample size calculation. Initial data screening identified 16 participants who had scored less than four on the Digit Span Backwards task, which indicated low effort and could result in low-quality data. As per the registered protocol these participants were excluded, and a further 16 participants were recruited to replace them. The final data set of 201 participants included 20 participants with a digit span score of less than 4 (including the 16 already identified) and one participant who failed the second attention check. These participants were excluded prior to analysis, so the final data set comprised of 180 participants.

Figure 3-1 Consort Diagram



3.2.6 Data Analysis

The main analysis consisted of creating a Pearson correlation matrix to evaluate whether the predictions about the correlations between tasks match the performance on these tasks in this sample. Equivalence testing was used to evaluate whether there was a difference in the strength of correlation of the correlation matrix. In addition, the internal consistency of the Bristol ERT and Dynamic ERT in this sample was assessed and used to calculate the Standard Error of Measurement (SEM) and the Smallest Real Difference (SRD). As the Bristol ERT includes low intensity trials that could be introducing random variability. A sensitivity analysis was conducted with a reduced version of the Bristol ERT, where only trials with stimuli presenting around 40% intensity or more were included. This means that for each participant the 36 lowest intensity trials were removed (6 trials per emotion), leaving a total of 60 trials.

The data was downloaded from Gorilla and loaded into RStudio (2020) R version 4.0.2 for data cleaning and analysis. The 'apaTables' package version 2.0.5 (Stanley, 2018) was used to calculate descriptive statistics and correlations for the relevant outcome variables. To assess whether there was evidence for a difference in strength of the correlations an r to z transformation method was used: http://quantpsy.org/corrtest/corrtest2.htm (Lee & Preacher, 2013; Steiger, 1980). The reliability analysis was conducted using the 'splithalf' package version 0.7.1 (Parsons, 2020) and *ltm* package version 1.1-1 (Rizopoulos, 2006).

3.3 Results

3.3.1 Demographics

180 participants were included in the final data set (64 female, 116 male). The average age was 27 years with a range of 18 to 67 years. Education data showed that 54% of participants reported having completed an undergraduate degree or higher, 22% completed secondary or further education, 13% indicated they had completed secondary school, and 1% of participants reported having completed vocational training. Participants from 28 different nationalities were included in the study (Appendix B, section 8.5). One participant reported that anxiety may influence their performance on the emotion recognition tasks, but this was not deemed a reason for exclusion.

3.3.2 Validity

The means and standard deviations for primary outcomes of the six measures included in this study and a correlation matrix of these outcomes are presented in Table 3-1. All the correlations are presented using Pearson's correlation coefficients.

Table 3-1

Means, standard deviations, and correlations of the tasks with confidence intervals

Variable	*M* (% acc.)	*SD*	1	2	3	4	5
1. Bristol ERT (total hits)	67.19 (70%)	7.51					
2. Dynamic ERT (total hits)	58.59 (61%)	9.27	.52** [.41, .62]				
3. TASIT-S: EET (total hits)	7.51 (75%)	1.44	.32** [.18, .45]	.27** [.13, .40]			
4. DSB (span length)	6.43	1.85	.10 [-.04, .25]	.05 [-.09, .20]	.06 [-.09, .20]		
5. GFMT (total hits)	32.76 (82%)	4.53	.29** [.15, .42]	.32** [.19, .45]	.10 [-.05, .24]	.07 [-.08, .21]	
6. TAS (total score)	49.56	11.9	-.13 [-.27, .02]	-.09 [-.24, .05]	-.08 [-.23, .06]	-.04 [-.18, .11]	-.14 [-.28, .01]

Note. M and *SD* are used to represent mean and standard deviation, respectively. Percentage accuracy is added in brackets. Values in square brackets indicate the 95% confidence interval for each correlation. The confidence interval is a plausible range of population correlations that could have caused the sample correlation (Cumming, 2014).
* indicates $p < .05$. ** indicates $p < .01$.

Means and standard deviations for performance on the Dynamic ERT, EET from TASIT-S, GFMT, and TAS are comparable to the normative data scores available for these tasks. For the Dynamic ERT adult scores have been reported between 51.9 and 66.9 with standard deviations of 7.7 to 12.1 (Kessels et al., 2014). Normative scores for the EET from TASIT-S in a US and Australian population were reported as *M* = 7.43 and *SD* = 1.42 (McDonald et al., 2018). Mean performance on the short version of the GFMT is 32.52 (81.3%) with a standard deviation of 9.7 (Burton et al., 2010). Normative data for the TAS collected in a healthy Canadian population showed the mean score on the scale was 45.75 with a standard deviation of 11.35 (Parker et al., 2003).

There was strong evidence for a positive correlation between the three emotion recognition tasks, with the strongest correlation observed between the Bristol ERT and the Dynamic ERT and a slightly lower correlation of both these tasks with the EET from TASIT-S. The difference between these correlations was evaluated using r to z transformation (Lee & Preacher, 2013; Steiger, 1980). Results indicated that the strength of the correlation between the Bristol ERT and Dynamic ERT was greater than the correlation between either of those and the EET from TASIT-S ($z = 2.55$, $p \leq .01$ and $z = 3.6$, $p \leq .01$ respectively). There was no evidence for a difference between the correlations of the Bristol ERT with the EET from TASIT-S compared to the Dynamic ERT with EET from TASIT-S ($z = 0.72$, $p = .47$), suggesting that the strength of these correlation was equivalent.

There was strong evidence for positive correlations of the Bristol ERT and the Dynamic ERT with the GFMT (measure of face discrimination). There was, however, no evidence for a correlation between performance on the EET from TASIT-S with the GFMT. The results indicated that none of the emotion recognition tasks were correlated with performance on the Digit Span Backwards (measure of working memory). To better evaluate the pattern of correlations, the difference in correlation between the emotion recognition tasks and the other tasks was also assessed using r to z transformation. For the Bristol ERT, the results indicated that the Bristol ERT was more strongly correlated with the Dynamic ERT than the GFMT ($z = 3.01$, $p \leq .01$), but there was no evidence for a difference in correlation between the Bristol ERT and the EET from TASIT-S compared to the Bristol ERT and the GFMT ($z = 0.32$, $p = .75$). There was also evidence that the Bristol ERT was more strongly correlated with both the Dynamic ERT and the EET from TASIT-S compared to the Digit Span Backwards ($z = 4.53$, $p \leq .01$ and $z = 2.23$, $p = 0.03$

respectively). Furthermore, there was weak evidence that the Bristol ERT had a stronger correlation with the GFMT than the Digit Span Backwards (z = 1.92, p = 0.05).

There was no evidence that the TAS was correlated with performance on any of the emotion recognition tasks, suggesting that alexithymia was not associated with emotion recognition performance in this sample. Exploratory analysis showed that according to the cut-off scores of high alexithymia ≥ 61 and low alexithymia ≤ 51 only 35 participants were classified as having high alexithymia, whilst 99 participants were classified as having low alexithymia. Welch's two sample t-tests were conducted to investigate potential group differences for high versus low alexithymia. There was no evidence for a group difference on any of the emotion recognition tasks in this sample, Bristol ERT ($t(57) = 0.88$, $p = 0.38$), Dynamic ERT ($t(50) = 0.33$, $p = 0.75$), and EET from TASIT-S ($t(58) = 0.42$, $p = 0.68$). The mean emotion recognition accuracy for the reduced Bristol ERT (low intensity trials removed) was 50.88 (SD = 4.84), which when converted to a percentage accuracy is slightly higher than the emotion recognition accuracy for the full Bristol ERT; 85% accuracy for the reduced version vs. 70% on the full version. The two versions of the Bristol ERT in this sample had a correlation coefficient of 0.92 ($p \leq .01$) and Confidence Interval of .90 to .94. The correlations of the reduced Bristol ERT with the other measures were approximately the same as for the full Bristol ERT (See Appendix C, section 8.6).

3.3.3 Reliability

Internal reliability of the Bristol ERT and Dynamic ERT was assessed using split-half reliability approach outlined by Parsons, Kruijt and Fox (2019) and by calculating Cronbach's alpha. This was not possible for the EET from TASIT-S as there were too few trials. Only trials for the 180 participants included in the validity analysis were used to estimate the reliability of the Bristol ERT and the Dynamic ERT in this sample. All the trials were treated as the same condition and 5000 random splits were used to establish a split-half reliability estimate and a Spearman-Brown corrected reliability estimate which corrects for task length. The confidence interval for Cronbach's alpha was also calculated using 5000 bootstrap samples.

For emotion recognition accuracy of the Bristol ERT, the Spearman-Brown corrected split half reliability estimate was $r_{sb} = 0.65$ (95% CI: 0.57 to 0.72). The split-half reliability estimate for the reduced Bristol ERT was only slightly higher at $r_{sb} = 0.68$ (95% CI: 0.61 to 0.75). Cronbach's alpha was $\alpha = 0.77$ (95% CI: 0.69 to 0.82) for the full Bristol

ERT and α = 0.75 (95% CI: 0.66 to 0.80) when only the 60 higher intensity trials were included. The Spearman-Brown corrected reliability estimate for emotion recognition accuracy on the Dynamic ERT was higher than both variations of the Bristol ERT at r_{sb} = 0.74 (95% CI: 0.69 to 0.79). Cronbach's alpha for the 60 trials on the Dynamic ERT was also higher α = 0.84 (95% CI: 0.80 to 0.87).

Standard Error of Measurement (SEM) and the Smallest Real Difference (SRD) provide useful information to interpret individual scores (See section 2.1.3 Interpreting reliability coefficients in Chapter 2 for a detailed description and method of calculation). The standard deviations and spearman brown corrected internal reliability estimates for used to calculate the SEM and SRD for both the Bristol ERT and the Dynamic ERT. The SEM for the Bristol ERT in this sample was 4.44 ($7.51\sqrt{1-0.65}$), which makes the SRD 12.3 hits (2.77×4.44). The SEM for the Dynamic ERT was 4.73 ($9.27\sqrt{1-0.74}$), which means the SRD based on this sample was 13.1 hits (2.77×4.73).

3.4 Discussion

The aim of the study was to provide evidence that the Bristol ERT is measuring the construct of emotion recognition, which in turn would allow for valid inferences to be made based on performance on the Bristol ERT. Performance on the Bristol ERT was compared with performance on two other emotion recognition tasks (Dynamic ERT and EET from TASIT-S), a working memory task (Digit Span Backwards), a face discrimination task (GFMT), and an alexithymia scale (TAS). A priori predictions about the correlations between these tasks were made and were used as a framework to link the results back to prior evidence and theory. Each hypothesis is discussed in turn to help evaluate whether the Bristol ERT is likely to be measuring emotion recognition.

3.4.1 Validity

The primary hypotheses stated that the three emotion recognition tasks would have a correlation coefficient of around 0.6 and that the Bristol ERT and Dynamic ERT would be more highly correlated with each other than with the EET from TASIT-S. The results broadly match the hypotheses made. The correlation coefficient for the Bristol ERT and Dynamic ERT was within range of the predicted estimate. The correlation of both those tasks with performance on the EET from TASIT-S was lower than expected but there was strong evidence for a correlation with the EET from TASIT-S. It is possible that the observed difference in strength of correlation was due to the type of stimuli used. The Bristol ERT and Dynamic ERT are based on emotion recognition from face stimuli, whilst vignettes used in the EET from TASIT-S include body language as well. This is supported by evidence that the GFMT is correlated with the Bristol ERT and the Dynamic ERT but not the EET from TASIT-S. Nevertheless, the lower correlation was surprising given the correlation of r = 0.67 between the EET from TASIT and an Ekman and Friesen stimuli emotion recognition test reported by McDonald and colleagues (2006). TASIT-S was developed as a screening tool for social cognition difficulties (Honan et al., 2016), so it is possible that the EET from TASIT-S might not be as sensitive to individual differences in emotion recognition in a healthy sample. It is worth considering that stimuli for the EET from TASIT-S include spoken scripts, so performance could be influenced by English language ability. Language was not explicitly controlled for in this study and could have influenced performance on the EET from TAST-S, although the scripts in the video clips are neutral and should not impact emotion recognition and performance in the current study was comparable to the normative data available.

The predicted pattern of correlation coefficients was that the three emotion recognition tasks would show the lowest correlation with the Digit Span Backwards, a slightly higher correlation with the GFMT and the highest correlation with each other. This pattern is clearly observed for the Bristol ERT compared to the Dynamic ERT, Digit Span Backwards, and GFMT. The equivalence testing indicated that there is a change in the strength of the observed correlations corresponding to the pattern of correlation coefficients. The Dynamic ERT presents in the same way as the Bristol ERT and as hypothesised, there is no evidence of a correlation between the three emotion recognition tasks and the Digit Span Backwards. As discussed, whilst the highest correlation of the EET from TASIT-S is with the other two tasks of emotion recognition the correlation is not as high as expected and appears to be equivalent to the correlation observed between those tasks and the GFMT. The difference in stimuli used could explain these results. Nevertheless, the lack of evidence for a correlation between the GFMT and the EET from TASIT-S was surprising given the correlation of the EET from TASIT with the Benton Face Recognition Task ($r = 0.45$) reported by McDonald and colleagues (2006).

Contrary to the final prediction made there was no evidence of a negative correlation with the TAS. It is also possible that the trend simply did not exist in this neurologically normal sample given that most participants scored low on the alexithymia scale. However, Rosenberg and colleagues (2019) did not find a correlation with the EET from TASIT and the TAS-20 in mixed Traumatic Brain Injury and neurologically normal sample. A possible explanation is that reaction times rather than accuracy on emotion labelling tasks such as the ones used in this study may be associated with alexithymia (Ihme, Sacher, Lichev, Rosenberg, Kugel, Rufer, Grabe, Pampel, Lepsien, Kersting, Villringer, Lane, et al., 2014; Ihme, Sacher, Lichev, Rosenberg, Kugel, Rufer, Grabe, Pampel, Lepsien, Kersting, Villringer, & Suslow, 2014) or that emotion recognition is not linked to or dependent on the ability to express emotion. Another explanation is that the tasks are not able to appropriately measure the constructs tested, although this seems unlikely given that other correlations were as predicted, and performance matches normative data for these tasks. The absence of predicted correlation for the TAS and weaker association with EET from TASIT-S are difficult to interpret in this type of validation study and require further investigation to better understand what is driving the effect or lack thereof (Streiner et al., 2015).

3.4.2 Reliability

The reliability estimates for scores on the Bristol ERT in this sample can be described borderline acceptable for a research context but would not be sufficient if a task were used in a clinical context to assess individual differences (Ponterotto & Ruckdeschel, 2007; Strauss et al., 2006). Notably, reliability coefficients associated with the Dynamic ERT in this sample were better. They would be considered adequate in a research setting although still too low to be considered suitable in clinical setting (Ponterotto & Ruckdeschel, 2007). In the context of this study, it means that the strength of correlation coefficients presented should be interpreted with caution because of the potential variance in the scores (Parsons et al., 2019). Estimates of internal consistency are influenced by the homogeneity if the construct being measured and number of items on a test (Ponterotto & Ruckdeschel, 2007; Streiner et al., 2015). It is possible that the reliability estimates in this study are decreased because the Bristol ERT and Dynamic ERT trials vary in intensity and type of emotion presented. Internal consistency reliability estimates for scores on the Bristol ERT increased when low intensity trials were removed. A decrease in trials would usually be associated with a decrease of the reliability estimates, as observed for Cronbach's alpha (Ponterotto & Ruckdeschel, 2007), which suggests that variance on low intensity trials was impacting split half reliability estimates. This means that lower split half reliability estimates should be expected for the Bristol ERT and are not necessarily indicative of low-quality data. This should be monitored and evaluated in future studies. It is also worth noting that low reliability estimates are not inherently problematic when considering group differences given a large enough sample size (Parsons et al., 2019; Streiner et al., 2015).

3.4.3 Strengths and Limitations

The current study compares performance on the Bristol ERT with a range of other tasks in a large-scale sample. Using equivalence testing to evaluate the correlations between the different tasks allowed for better inferences about the pattern of correlations observed (Lakens et al., 2020). A limitation of this study was that the EET from TASIT-S may not have been appropriate to measure individual variation in a healthy population, which restricts the comparison with the other two emotion recognition tasks. Unfortunately, due to time constraints it was not possible to use the full-length EET from TASIT, which would probably have allowed for a better comparison. Furthermore, the Digit Span Backwards may not have been an appropriate measure of

working memory for an online study. It is possible that participants wrote down digits presented instead of trying to keep them in mind, although mean performance in this study was comparable to performance reported by Hilbert and colleagues (2015). It is also worth noting that participants signed up to Prolific are a self-selected sample, who are actively interested in contributing to research and are generally well educated (Peer et al., 2017). This could impact generalisability of results.

3.4.4 Implications and future directions

Overall, the findings indicate that the Bristol ERT is measuring a shared underlying construct with the other two emotion recognition tasks, which is assumed to be the ability to recognise emotions. Performance on the Bristol ERT seems to be in part based on face recognition but is unlikely capturing general cognitive ability because there is no evidence for a correlation with the working memory task. The fact that the results for the Bristol ERT largely correspond to the a priori predictions made provides support for the validity of inferences about emotion recognition made using performance on the Bristol ERT in a neurologically normal sample. Further research should aim to better understand the findings that were contrary to predictions made in this study and continue the validation of the Bristol ERT. How well a task can discriminate between two groups that are thought to differ on a given construct is form of construct validation (Streiner et al., 2015). Given current evidence indicating that emotion recognition is impaired after TBI (Babbage et al., 2011) the ability to replicate this finding using the Bristol ERT would provide further evidence for the validity inferences about emotion recognition made using this task.

Chapter 4 Assessing emotion recognition after moderate to severe Traumatic Brain Injury using the Bristol Emotion Recognition Task

4.1 Introduction

The first two studies presented in this thesis investigated psychometric properties of the Bristol Emotion Recognition Task (ERT) in neurologically healthy populations. To evaluate whether the Bristol ERT could be suitable to assess emotion recognition after TBI the performance on this task needed to be investigated in this population. As was outlined in the introductory chapter of this thesis, emotion recognition is not commonly assessed in a clinical setting (Kelly et al., 2017). The Bristol ERT is available as part of the Cambridge Automated Neuropsychological Test Battery, which means it could easily be used for neuropsychological assessment. To help evaluate the utility of the Bristol ERT it is important to assess whether it is possible to discriminate between participants with TBI and neurologically healthy controls based on task performance. Furthermore, associations between emotion recognition and potentially confounding factors need to be understood to allow for appropriate interpretation of scores. A feasibility study using the Bristol ERT in an acquired brain injury population was conducted as part of a masters project (Müller-Glodde, 2015). The current study builds on that work by using the Bristol ERT to assess emotion recognition in a larger sample of participants with moderate to severe TBI in a clinical setting. The main aim was to check whether the Bristol ERT could detect changes in emotion recognition in a moderate to severe TBI population that are similar to those reported in the current literature. The study presented in this chapter was registered on the Open Science Framework and sections are taken from the pre-registered protocol (10.17605/OSF.IO/EXS39).

4.1.1 Moderate to severe TBI and Emotion Recognition

There is extensive evidence that people with moderate to severe TBI perform worse than neurologically healthy controls on emotion recognition tasks (Babbage et al., 2011; Murphy et al., 2021). Babbage and colleagues (2011) conducted a meta-analysis of studies using static facial stimuli to investigate emotion recognition accuracy after moderate to severe TBI. They found that participants with moderate to severe TBI consistently performed lower than non-brain injury controls and reported that the effect size for this deficit was large (Hedge's g = 1.1, 95% CI 0.97 to 1.25). Ietswaart and colleagues (2008) showed that emotion recognition difficulties in a TBI population

persist one year post injury. They also argued that the decrease in emotion recognition associated with TBI is directly related to the brain injury and cannot be attributed to confounding factors or differences in recovery process, as they used people with orthopaedic injury as a control group.

4.1.1.1 Emotion specific differences in emotion recognition accuracy

The individual emotions associated with decreased performance for the TBI group differs across studies, but generally include anger and disgust. Rosenberg and colleagues (2014) reported that participants with TBI scored lower than controls on anger, disgust, and happiness, and observed a trend towards lower scores for the other emotions. Using the same task Rosenberg and colleagues (2015) reported that participants with TBI scored lower for anger, disgust, and fear. Rigon and colleagues (2016) found that participants with TBI scored lower only for anger and disgust on a static emotion recognition task, and then anger, disgust, and sadness on the dynamic ERT used by Rosenberg and colleagues (2015; 2014). Meanwhile, a recent meta-analysis indicated that participants with moderate to severe TBI were impaired across all emotions, although the size of the effect was smaller for positive emotions (happiness, surprise) than negative emotions (anger, disgust, sadness, fear; Murphy et al., 2021). Studies commonly report that participants with TBI show impairments predominantly for negative emotions on emotion recognition tasks (Babbage et al., 2011; Croker & McDonald, 2005; Hopkins et al., 2002). Rosenberg and colleagues (2014) investigated this valance effect. They found that participants with moderate to severe TBI were more impaired on negative compared to positive emotions, but that this result was affected by the intensity of the emotion presented. When levels of intensity were adjusted relative to emotion difficulty, participants with moderate to severe TBI showed an impairment across all six 'basic' emotions (Rosenberg et al., 2015). This suggests that increased impairment for negative compared to positive emotions is driven by differences in the difficulty with which individual emotions are identified (Adolphs, 2002b; Rosenberg et al., 2014). Differences in task demands and analysis techniques are likely contributing to the inconsistent results regarding individual emotion performance.

4.1.2 Observational research: Cross-sectional studies

Observational studies are used to study TBI because it is simply put not ethical to randomly give people a TBI in order to conduct a Randomised Control Trial. Cross sectional research studies are a type of observational study that can be used to study

associations between TBI and other factors of interest (Mann, 2003). They are a quick and easy way of establishing associations between variables of interest and can be used to inform the direction and design of future studies (Mann, 2003). A disadvantage of the research design used in this study is that it was not possible to infer causality because there was no way to establish a temporal sequence for the variables. Potentially confounding factors linked to emotion recognition and TBI need to be considered and controlled for where possible. The impact that selection bias based on the method of recruitment could be having should also be considered (Mann, 2003).

4.1.3 Current study: Rationale

To evaluate whether the Bristol ERT could be useful in a neuropsychology setting it is important to conduct research in that setting. Participants with moderate to severe TBI for the current study were recruited from an outpatient head injury rehabilitation service as this is a clinical setting in which the Bristol ERT could prove useful. The purpose was to assess whether the participants with moderate to severe TBI scored lower than neurologically healthy controls on overall emotion recognition accuracy (total). This finding would indicate that the Bristol ERT is comparable to other emotion recognition tasks used in a TBI population. The Bristol ERT could also provide useful information about individual emotion recognition accuracy and bias. The studies investigating individual emotion recognition accuracy have largely used hits (total number of correct responses) as an outcome measure and do not consider the misidentifications made (Croker & McDonald, 2005; Rosenberg et al., 2015; Rosenberg et al., 2014). This is potentially problematic because there is evidence that people with TBI do not have the same response patterns to healthy controls (Callahan et al., 2011). Individual emotion recognition accuracy in this study was assessed using an unbiased hit rate (See introduction section 1.3.2 for a detailed description of the Bristol ERT). This may give a better understanding of emotion specific difficulties experienced by people with moderate to severe TBI. For example, the feasibility study conducted using the Bristol ERT in a brain injury population indicated that emotion recognition accuracy was reduced across all emotions other than happiness, and it is possible that the sample was simply not big enough to detect a difference in performance on happiness (Müller-Glodde, 2015). There was also evidence that anxiety was associated with an increased likelihood of misidentifying faces as fearful and decreased likelihood of misidentifying them as happy. Given that there is an increased prevalence of anxiety after TBI (Osborn

et al., 2016), it is possible that anxiety is moderating some of the differences in emotion recognition associated with having a TBI.

4.1.3.1 Anxiety as a potential moderator

Having an anxiety disorder has been associated with decreased overall emotion recognition (Plana et al., 2014) and state anxiety has been shown to be associated with decreased emotion recognition on the Bristol ERT (Attwood et al., 2017; Dyer et al., 2022). Assessing whether anxiety is associated with performance on the Bristol ERT in a TBI population could help with the interpretation of scores indicating emotion recognition difficulties in this population. Especially the presence of negative attribution bias could be associated with anxiety as opposed to the TBI. Richards and colleagues (2002) reported that people with high levels of anxiety were more likely to identify ambiguous stimuli as fearful and when anxiety was induced participants in the anxiety condition were more likely to identify faces as angry and less likely to identify them as happy. Attwood and colleagues (Attwood et al., 2017) used a variation of the Bristol ERT called the Bristol Emotion Bias Task (EBT) to investigate attribution bias towards angry faces. They found evidence that participants in the induced anxiety group were more likely to identify ambiguous stimuli as angry. Consequently, two versions of the Bristol EBT were also included in this study (see section 4.2.3.4 in Method for further details).

4.1.3.2 Confounding factors

To appropriately interpret and address difficulties that people with moderate to severe TBI are presenting with, it is vital to understand associations between emotion recognition and potentially confounding factors. Demographic factors, including, age (Byom et al., 2019) and sex (Thompson & Voyer, 2014) are associated with changes in performance on emotion recognition tasks and will be considered as confounders in the current study (See introductory chapter section 1.4.2 for details). Further, higher levels of education have been associated with better outcomes after TBI (Mushkudiani et al., 2007), so education will also be considered as a potential confounder.

There is evidence that emotion recognition tasks are impacted by memory, working memory, and processing speed (Yim et al., 2013), although general cognitive ability does not fully explain the decrease in performance on emotion recognition tasks observed in a TBI population (Rosenberg et al., 2015). Given the short presentation time of stimuli on the Bristol ERT it is possible that the task is particularly vulnerable to

difficulties with processing speed and attention. Consequently, information about cognitive ability from neuropsychological assessment will be collected as part of this study and explored as a confounder. Furthermore, participants will also be asked to complete a face discrimination task to check whether face perception ability can explain performance on the Bristol ERT (J. D. Henry et al., 2015). Williams and Wood (2010) have argued that alexithymia could underlie the difficulties in emotion recognition observed after TBI. It is possible these difficulties are particularly pronounced when facial expressions are only presented for a short time (Ihme, Sacher, Lichev, Rosenberg, Kugel, Rufer, Grabe, Pampel, Lepsien, Kersting, Villringer, Lane, et al., 2014). This highlights the importance of considering alexithymia as a confounder when assessing emotion recognition in a TBI population using the Bristol ERT. As discussed in the introduction of this thesis (Section 1.4.2) other factors, such as aggressive tendencies (Hoaken et al., 2007; Neumann et al., 2017) and depression (Dalili et al., 2015) are also associated with changes in emotion recognition and could be impacting performance on emotion recognition tasks after TBI. These factors were not considered as confounders in this study, but measures were included for exploratory purposes.

4.1.4 Current study: Aims and Hypotheses

The main aim of the current study was to investigate the association between emotion recognition and moderate to severe TBI using the Bristol ERT. Replication of findings indicating decreased emotion recognition in the TBI population compared with controls would indicate that the Bristol ERT could be useful to assess emotion recognition in a neuropsychology setting. Consequently, the primary hypothesis for this study was:

H1 Participants with TBI will be less accurate at recognising emotions on the Bristol ERT than neurologically healthy controls.

Participants with brain injury scored lower than non-injury controls on all emotions other than happiness in the feasibility study, and there was a trend towards lower scores on happiness (Müller-Glodde, 2015). As this study consisted of a bigger sample, the prediction for individual emotion recognition performance was that participants with TBI would score lower on all six emotions included on the Bristol ERT.

H2 Participants with TBI will present with decreased emotion recognition accuracy across all six 'basic' emotions

The secondary aim of this study was to investigate anxiety as a potential moderator of performance on emotion recognition tasks following TBI. Our prediction was that anxiety would be associated both with emotion recognition accuracy on the Bristol ERT and attribution bias on the Bristol EBT.

H3 Anxiety will be associated with performance on the Bristol ERT after presence of TBI has been taken into consideration

H4 Participants with high levels of anxiety will show a bias towards perceiving angry and sad faces on the Bristol EBT

Finally, potentially confounding factors for the association between emotion recognition and TBI will be assessed to gain a better understanding of factors that could be impacting performance on the Bristol ERT. Factors considered in this study were anxiety, sex, education, face perception, and alexithymia. Exploratory analysis of the impact that depression, stress, and aggression could have on emotion recognition performance after TBI was also conducted.

4.2 Method

4.2.1 Design

The study was an observational, cross-sectional study investigating the association between having a TBI, and anxiety at the time of testing with emotion recognition from facial expressions. Emotion recognition performance was measured using a six-alternative forced-choice emotion recognition task (Bristol ERT). The primary outcome measure was overall emotion recognition accuracy on this task. Secondary emotion recognition outcomes included emotion specific emotion recognition accuracy and a measure of response bias for each emotion. The Depression, Anxiety, and Stress Scale (DASS-42) was completed at the time of testing and used as a measure of self-reported anxiety. For a full description of tasks and questionnaires completed see measures and materials section. The study has been reviewed by the South West – Central Bristol Research Ethics Committee and approval obtained from the Heath Research Authority and Health and Care Research Wales (19/SW/0062).

4.2.2 Participants

The TBI group were recruited from patients accessing out-patient head injury services due to a moderate to severe TBI. Patients who were deemed appropriate for the study by a clinician (based on the inclusion/exclusion criteria below) were approached about the study and asked to complete a consent form if they were interested in taking part.

After a patient had completed the study, an age and sex matched control participant was recruited online through Prolific (https://www.prolific.co). Screening criteria on Prolific were set such that only individuals meeting the following criteria were invited to take part: 1) resident in the UK, 2) same sex as the patient, 3) within a 5-year age range of the patient, 4) reported that they had not had a TBI or any other neurological condition, and 5) matched the additional inclusion/exclusion criteria below. See Appendix A for this chapter (Section 8.7) for a copy of the background questionnaire given to participants.

Patient participants took part in the study on a voluntary basis and were not reimbursed for their time but were asked whether they would like to be contacted with the outcomes of the study. Matched control participants were reimbursed £0.50 to complete the screening task and if eligible, a further £6 for completing the full study.

4.2.2.1 Inclusion criteria

Patients

- Attending services due to moderate to severe TBI
- Above the age of 18 and deemed to have capacity
- Basic English language skills, which were deemed sufficient to understand and complete the tasks as assessed by the responsible clinician

Matched control participants

- Same age and sex as a patient who completed the study
- Above the age of 18 and deemed to have capacity
- Basic English language skills, which were deemed sufficient to understand and complete the tasks

4.2.2.2 Exclusion criteria

Patients

- Visual and perceptual deficits
- Insufficient cognitive ability, or significant attentional deficits and fatigue
- Presenting with both language and physical/motor difficulties
- Fluctuating capacity (Capacity was assumed, as clinicians were asked not to approach patients if they were unsure whether capacity would remain stable)

Matched control participants

- History of brain injury
- Visual and perceptual deficits
- Insufficient cognitive ability, or significant attentional deficits and fatigue
- Presenting language difficulties
- Presenting physical motor difficulties
- Fluctuating capacity (Capacity was assumed)

4.2.2.3 Control group amendment

The original study design included a group of control participants called associated controls, which consisted of family members and friends of the patients taking part in the study. Due to recruitment difficulties, it was not feasible to use this group as a control. As a contingency, a group of age and sex matched controls were recruited online through Prolific.

4.2.2.4 Sample size determination

The sample size was originally proposed as 120 total (60 patients, and 60 controls), but these targets were adjusted due to COVID-19 related delays, and the rate of recruitment based on the number of patients with TBI attending Neuropsychology services within the time frame of the study. The updated sample size used was 52 participants in total for this study, 26 in the patient group, and 26 in the matched control group.

Both the original and adjusted sample size were based on an *a priori* sample size calculation conducted in G*Power 3.1 (Faul et al., 2007). The primary analysis proposed (see Data Analysis section) was a linear multiple regression with 7 predictors, of which only TBI was considered a tested predictor for the sample size calculation. The effect size for the difference in overall emotion between the brain injury and control groups was d = 1.22 in the pilot study (Unpublished: Müller-Glodde, 2015), which is comparable the effect size reported in the meta-analysis by Babbage et al. (2011). However, effect sizes from small sample studies are often inflated (Button et al., 2013), so the effect size from the pilot study was reduced by one third to a value of d = 0.8 and then converted into an f^2 value of 0.16. Based on that effect size and the stated number of predictors a sample size of 115 was required to obtain 95% power at an alpha level of 1%. This sample size was not reached, but our achieved sample size of 52 was sufficient to obtain 80% power at an alpha level of 5%.

4.2.3 Measures

4.2.3.1 Questionnaire - Background information

Participants were given a short questionnaire to complete including some demographic questions (such as sex, age, years of education, and occupation) and some broad questions about brain injury. For example, asking whether they have had a brain injury, and/or other neurological disorder if in the control group; or how long since their TBI, and whether they have had repeated TBI if in the patient group (See Appendix A, section 8.7).

4.2.3.2 Screening task

A short screening task was presented to participants at the start of the testing session to check whether they understood the task instructions and could cope with the

demands of the task. They were asked to identify six different types of fruit and vegetables in the same way they would have to identify emotions during the Bristol ERT. Each fruit/vegetable was presented three times, resulting in 18 trials total. The task took no more than two minutes to complete and provided participants an opportunity to get used to the experimental set up. Furthermore, participants who scored less than 80% on the task were excluded from the study, as they were unlikely to be able to complete the Bristol ERT in a meaningful way if they could not accurately complete the screening task.

4.2.3.3 Bristol Emotion Recognition Task (Bristol ERT)

Participants were prompted to take a break after half of the trials had been completed, and once they had responded to all 96 trials a text screen informed them that they had completed the task. The primary outcome was 'total hits' (total number of correct identifications), used as a measure of overall emotion recognition accuracy. In addition, 'hits' (number of times an emotion was correctly chosen) and 'false alarms' (number of times an emotion was erroneously chosen) were recorded for each emotion. The hits and false alarms were used to calculate an 'unbiased hit rate' (H_u) for each emotion, calculated by dividing the number of hits squared by 16 times the sum of hits and false alarms for that emotion). Additionally, a response 'bias score' was calculated by adding the number or hits and false alarms for each emotion to give the number of times an emotion was selected as a response.

4.2.3.4 The Bristol Emotion Bias Task (EBT) – Happy - Anger & Happy - Sadness

The EBT tasks are variation of the Bristol ERT designed to investigate bias. Instead of creating morph sequences using a prototypical face and a target emotion, a 15-image morph sequence was created using two target emotions. In this case, images of 'happy' expressions were merged with either 'sad' or 'angry' expressions to create a series of ambiguous stimuli. This means that each image consisted of a certain percentage of 'happy' and certain percentage of either 'sad' or 'angry' expressions, the only exceptions being the first and the last image which were 100% intensity images of the two emotions. The images were presented in the same way as during the ERT, but participants had only two response options and were asked to label the expression presented as either 'happy' or 'sad/angry'. Each image was presented three times and the two morph sequences used for this study were presented consecutively, first the happy/angry EBT and then the happy/sad EBT. This means a total of 90 images were

presented in two blocks of 45 and the presentation of the images within each block was randomised. As with the Bristol ERT, a text screen appeared at the end of each section informing participants that they had completed that stage and gave instructions about how to proceed. There was a break between the two EBT blocks. The outcome measure was a balance point representing the number of 'happy' faces perceived as a proportion of the total number of trials. It was calculated by dividing the number of times a face was labelled as 'happy' by the total number of trials (45) and then multiplying this number by 15 (the number of images in the morph sequence). A higher balance point indicates an increased tendency to label faces as happy, so a score of 0 means that all faces were labelled as sad and a score of 15 suggests all faces were labelled as happy.

4.2.3.5 Glasgow Face Matching Test (GFMT) – short version

The GFMT assesses people's ability to recognise and discriminate between unfamiliar faces. The short version was used in this study, meaning that 40 pairs of faces were presented, and participants were asked to make a judgment as to whether the two pictures presented were of the same person or two different people. The photographs were presented side-by-side and remained visible on the screen until a response was given. 20 trials showed faces of the same person and 20 showed faces of two different people. A screen at the end informed participants that they had completed the task. Total hits (number of correct responses) was used as the outcome measure for this study. Normative data for the GFMT is available (Burton et al., 2010).

4.2.3.6 Depression, Anxiety, and Stress Scale (DASS-42)

The DASS is a self-report questionnaire commonly used to assess mood in clinical practice but also suitable for research purposes. There are 42 items, scored for three subscales of 14 items each measuring depression, anxiety, and stress levels. The items are rated on a scale from 0 to 3, where 0 = Did not apply to me at all and 3 = Applied to me very much, or most of the time. The maximum score on the DASS is 126, with each subscale having a maximum score of 42 (14 items with a maximum score of 3 on each item). The scores on each subscale as well as overall score were used as continuous outcome measures in this study. The task has been validated (Lovibond & Lovibond, 1995) and normative data as well as cut-off scores for the DASS are available for a UK based sample (Crawford & Henry, 2003).

4.2.3.7 Toronto Alexithymia Scale (TAS-20)

The 20 item TAS was used in this study, which can be separated into three subscales (Bagby, Parker, et al., 1994; Bagby, Taylor, et al., 1994). These include

1) Difficulty in identifying feelings, seven items
2) Difficulty in describing feelings, five items
3) Externally orientated thinking, eight items.

The aim was to capture both affective processing and cognitive strategies that contribute to the construct of alexithymia (Bagby & Taylor, 1997). Items were rated using a 5-point Likert scale whereby 1 = strongly disagree and 5 = strongly agree. This means the minimum score was 20 and the maximum score was 100, as 5 of the items are negatively keyed and are reverse scored when calculating the total score. The TAS-20 was developed to be used as a continuous variable, but cut-off scores are available, non-alexithymia ≤ 51, possible alexithymia 52 to 60, and alexithymia ≥ 61 (Bagby & Taylor, 1997). The total score was used as a continuous outcome measure in this study. Normative data from a Canadian population is available from Parker et al. (2003). An attention check was added halfway through, asking participants to give a response of "4" for that item.

4.2.3.8 Buss-Perry Agression Questionnaire (BPAQ)

The BPAQ consists of 29 items that load onto four subscales. The subscales are physical aggression (9 items), verbal aggression (5 items), hostility (8 items), and anger (7 items). Participants were asked to rate items using a 5-point scale, whereby 1 = extremely uncharacteristic of me and 5 = extremely characteristic of me. Two of the items are reverse scored, so the maximum score was 145. Higher scores indicate higher tendency towards aggression and the total score was used as a continuous outcome measure in this study. Normative data is available from Buss and Perry (1992).

4.2.4 Procedure

Data collection for this study was disrupted due to COVID and the study had to be moved online part way through to be completed. At the same time the control group was switched from associated controls to matched controls. The procedure for in person data collection is described first followed by a section outlining the changes made to move the study online and recruitment for matched controls.

4.2.4.1 Before the testing session

After being approached by a clinician and agreeing to take part in the study patient participants were given a consent form to sign and study booklet to complete. The study booklet summarised information about the study and contained the background questionnaire, TAS and BPAQ. Participants were asked to return their signed consent form and completed questionnaires at a routine appointment prior to the testing session or at the beginning of their testing session. The testing session was arranged with the research team before or after a routine clinical appointment. If the questionnaires were not completed before the testing session, participants were asked to complete the forms at the end of the testing session.

4.2.4.2 Testing session

The testing took place at the study site in clinical interview rooms. Before starting the tasks, participants were given an opportunity to ask questions and asked to verbally reconfirm consent. Participants completed the screening task, the Bristol ERT, EBT, and GFMT on a dedicated laptop. Presentation of the experimental tasks and collection of response data was managed through EPrime software (Schneider et al., 2002). As the tasks had no obvious right or wrong answer, researcher bias through presence in the testing room was considered to be low. There were opportunities to take breaks between each task, as well as scheduled breaks during the Bristol ERT. There was no time limit for the breaks and patients were given the opportunity to go for a walk if required. The DASS was presented as paper-and-pen questionnaire once the experimental tasks were completed. At the end of the session participants were given a verbal debrief and asked to confirm that they were happy for their data to be used, after which their participation in the study was completed.

4.2.4.3 Post-testing session

Responses on the DASS and other questionnaires were scored and recorded electronically by the research team. A copy of the DASS was added to the patient medical file and the clinical team made aware of the scores. In conjunction to the above measures, data was collected from patient medical files. The information gathered included information about their TBI (e.g. time since injury, how the injury occurred, and information about severity of injury), as well as cognitive test scores from the Wechsler Adult Intelligence Scale (WAIS) VI or Repeatable Battery for the Assessment of Neuropsychological Status (RBANS) if available.

4.2.4.4 Online testing (COVID-19 adjustments and matched controls)

When data collection resumed after COVID, not all patients attended appointments in person, so clinicians were asked to approach patients and verbally complete a consent to be contacted form if patients were interested in taking part (Appendix B, section 8.8). All participants were required to have access to a tablet or computer with internet connection to take part in the study. The consent form and study documents were sent to potential participants using the email provided on the consent to contact form and they were invited to ask questions at their next appointment or by emailing the research team. The consent form could be returned via email or in person at which point an online testing session was arranged. A secure online meeting platform called Attend Anywhere was used for the testing session. At the beginning of the call the patient participants were invited to ask any questions and given on overview of the testing session. They were then given a participant ID that allowed them to access an online version of the study and asked to complete the tasks and questionnaires whilst on a call to the researcher. This meant that like during an in person testing session they had opportunities to ask questions and the researcher could monitor the session in case there were any issues. Same as for the in-person session, participants were given a verbal debrief at the end of the session and informed that this would end their participation. Matched controls were all recruited online through Prolific (as described above in section 4.2.2).

The screening task for Prolific participants and the study tasks were all presented using Gorilla Experiment Builder (https://app.gorilla.sc; Anwyl-Irvine et al., 2019). Both patient participants and matched controls were given the option to view the information sheet again and asked to complete an online version of the consent form before they were given access to the tasks and questionnaires. Patient participants first completed the background questionnaire, then the four experimental tasks (Screening task, Bristol ERT, EBTs, and GFMT), and finally the three questionnaires (DASS, TAS, and BPAQ). The matched controls had already completed the background questionnaire and were directed straight to the other tasks and questionnaires. There were opportunities to take breaks between each task, as well as a scheduled break during the Bristol ERT. An attention check was added to the TAS asking participants to respond with 4 for to check that they were following instructions given. At the end of the testing session participants were asked to reconfirm consent and give permission for their data to be used as part of this study.

4.2.5 Data Analysis

The data was loaded into RStudio (2020) R version 4.0.4 for data cleaning and analysis. The main analysis packages used were the core *stats* package in R, *psych* package version 2.1.9, *apaTables* package version 2.0.5 (Stanley, 2018), *olsrr* package version 0.5.3 (Hebbali, 2017), *splithalf* package version 0.7.2 (Parsons, 2020), and *ltm* package version 1.1-1 (Rizopoulos, 2006).

4.2.5.1 Data checks

The primary outcomes for each task and questionnaire were checked to make sure that they fell within the expected range. Participants were removed prior to analysis if they scored less than 80% on the screening task. Outliers on the Bristol ERT were identified by checking whether any scores were more than 1.5 times the inter quartile range below the 1st quartile or above the 3rd quartile. Participants who failed the attention check during the online study, or who were considered outliers on the Bristol ERT were identified and a sensitivity analysis was conducted by repeating the primary analysis with these participants removed.

4.2.5.2 Primary analysis

Multiple linear regression was used to investigate the association between TBI and overall performance accuracy on the ERT (total hit rate). The unadjusted model included only TBI as a predictor of emotion recognition performance, where TBI was coded as a binary variable (patients with TBI vs matched controls). The second model additionally included the score for the anxiety subscale of the DASS to investigate the association between anxiety and emotion recognition performance. Finally, the model was adjusted for demographic factors, face matching, and emotion labelling. This means the final model included 6 variables: TBI, DASS anxiety score, sex (binary), years of education (continuous), GFMT score, and TAS score. The pre-registered protocol stated that age would also be controlled for but given that the control group was matched for age this was no longer deemed necessary.

4.2.5.3 Secondary analyses

To check whether there were emotion specific changes in emotion recognition after TBI the above regression analysis was repeated using the unbiased hit rate and bias scores on the Bristol ERT as emotion specific outcomes. The aim was to check for

potential emotion recognition effects that were hidden when only considering overall performance accuracy. Using the bias score to investigate emotion specific performance was not included in the pre-registered protocol but was added to provide information about response bias. To investigate how factors included in this study are associated with bias in facial affect recognition, the regression analysis was also run using the EBT outcomes. The association of anxiety with performance on the EBT was of particular interest. An exploratory analysis was conducted to investigate whether overall DASS score, DASS depression, DASS stress, or BPAQ score were associated with a change in overall emotion recognition after TBI. Models 2 and 3 of the primary analysis were rerun substituting the DASS anxiety score with each of those factors in turn.

4.2.5.4 Other analyses – deviations from the pre-registered protocol

The protocol stated that associations between severity of injury and general cognitive ability on emotion recognition performance in the TBI group would be explored. These analyses were not conducted as all the patient participants had moderate to serve TBI and there was insufficient information about general cognitive ability in the patient notes to allow for a meaningful analysis using the measures collected. A preliminary analysis of the mediatory effect of anxiety on emotion recognition after TBI was also outlined in the protocol but was deemed unnecessary given the results. The comparison between patient participants and their associated controls was not completed due to low recruitment of associated controls.

There were also several analyses conducted that were not included in the protocol. First, a sensitivity analysis was conducted excluding participants who failed the attention check for the online study by repeating the primary regression analysis those participants and their corresponding matched controls removed. Second, an ANOVA was conducted to check whether the modality of the testing session influenced performance on the Bristol ERT. The aim was to check for differences in performance between the participants with TBI who completed an in-person session and those who completed an online session. Finally, the reliability of performance on the Bristol ERT in this study was investigated using split half reliability and Cronbach's alpha. Both overall reliability and reliability for the TBI group and the matched control group separately were calculated.

4.3 Results

4.3.1 Descriptive statistics

A total of 52 participants were included in this study, 26 patients with TBI and 26 age and sex matched controls. None of the participants failed the screening task and so they were all included in the analysis. There was a higher proportion of males in this sample, 34 males (65%) and 18 females (35%). Mean age for the two groups was 46.7 years (SD = 16.2) for the TBI group and 47 years (SD = 15.8) for the control group. The age at which participants left education was comparable across the two groups, at 19.4 years (5.3) for the TBI group and 19.5 years (2.9) for the control group ($F(1, 50) = 0.004$ $p = 0.95$). Descriptive statistics and a comparison of performance across groups using an ANOVA are presented in Table 4-1.

Table 4-1

Descriptive statistics for questionnaires and cognitive tasks

Variable	TBI n = 26				Controls n = 26				Group comparison	
Questionnaires	**Mean**	**SD**	**Min**	**Max**	**Mean**	**SD**	**Min**	**Max**	***F-value***	***p***
DASS Total	31.0	23	1	84	28.1	22	0	80	0.21	0.65
Depression	10.6	9.2	0	34	11.2	10	0	32	0.05	0.83
Anxiety	8.27	7.2	0	27	4.81	5.3	0	21	3.90	0.05
Stress	12.2	8.8	0	32	12.2	10	0	34	0.00	1.00
TAS Total *	53.0	9.8	34	78	46.1	12	26	68	5.04	0.03
BPAQ Total *	70.2	19	34	115	67.9	22	40	113	0.15	0.70
Cognitive tasks	**Mean**	**SD**	**Min**	**Max**	**Mean**	**SD**	**Min**	**Max**	***F-value***	***p***
Bristol ERT *(total hits)*	52.9	8.6	39	70	64.2	8.8	45	78	22.2	<.001
EBT Angry *(balance point)*	7.31	1.5	2.3	10.3	6.85	1.6	3.7	10.7	2.07	0.16
EBT Sad *(balance point)*	6.31	1.7	0.7	8.7	6.91	1.3	3.7	10.0	1.11	0.30
GFMT *(total hits)*	30.0	5.0	20	38	32.2	4.8	17	39	2.68	0.11

** One patient did not complete TAS and BPAQ so the number of participants in the TBI group on those two scores is 25 instead of 26.*

Note: TBI is Traumatic Brain Injury. DASS is Depression, Anxiety, and Stress Scale. TAS is Toronto Alexithymia Scale. BPAQ is Buss-Perry Aggression Questionnaire. EBT is Emotion Bias Task. GFMT is Glasgow Face Matching Test. SD is the standard deviation. Group comparisons conducted using analysis of variance

Results showed that there was weak evidence for increased anxiety and higher alexithymia scores in the TBI group. Subsequent analysis of the three subscale on the Toronto Alexithymia scale, showed that there was evidence for a difference in performance only on the Identifying Feelings subscale ($F(1, 50) = 5.45$ $p = 0.02$). There was strong evidence for a difference in performance between the groups on the Bristol ERT but not the other cognitive tasks. The correlation between the variables was investigated using a correlation matrix (Appendix C, section 8.9). There was evidence that performance on the Bristol ERT was correlated with performance on the GFMT and for a correlation between years of education and TAS score. There was also evidence for positive correlations between nearly all the questionnaire outcomes, suggesting that they may be capturing a shared underlying factor. Additionally, all the DASS outcomes were positively correlated with the balance point on the sad EBT. This means participants who scored higher on the DASS identified more faces as happy on the Happy – Sad EBT.

4.3.2 Traumatic Brain Injury information

All the participants in the TBI group experienced a moderate to severe TBI and were being seen by Neuropsychological services at the time of participation. Mean time since injury in the TBI group is 43.5 months with a range of 2 to 288 months. Cognitive performance scores for the TBI group are presented in Table 4-2. Scores suggest that the participants in this study had decreased cognitive abilities. Unfortunately, the measures completed were not consistent within the patient group, so it was not possible to conduct an analysis including these scores that would allow for a meaningful interpretation of results.

Table 4-2

Summary information about TBI group cognitive testing scores

Cognitive indicator	n	Mean	SD	Min	Max
WAIS Working Memory Index	12	100	12	83	122
WAIS Processing Speed Index	14	89	15	65	105
RBANS Immediate Memory	13	83	19	49	126
RBANS Delayed Memory	13	82	19	44	101
RBANS Attention	11	82	19	43	115

Note: WAIS is the Wechsler Adult Intelligence Scale and RBANS is the Repeatable Battery for the Assessment of Neuropsychological Status. SD is standard deviation.

Of the 26 participants with TBI who took part in this study, 15 completed the study during an in-person testing session and 11 completed the study online whilst on a video call with the researcher. Mean performance on the Bristol ERT for the participants who attended in person was 52.5 (8.9) and 53.5 (8.7) for the participants who completed the study online. An ANOVA was used to check whether there was evidence for a difference in performance based on how the task was completed. Results showed no evidence for a difference in performance between participants with TBI who attended in person versus online ($F(1, 25) = 0.08$, $p = 0.78$).

4.3.3 Primary analysis

Multiple linear regression with overall emotion recognition measured using the Bristol ERT (total hits) as the outcome showed evidence for decreased emotion recognition in the TBI group compared to controls (Model 1, Table 4-3). When the model was adjusted for anxiety by including the DASS anxiety score there was still evidence for a decrease in performance between the TBI and control group (Model 2, Table 4-3). There was no evidence that anxiety was associated with emotion recognition performance in this sample. Finally, the model was adjusted for sex, age that participants left education, face discrimination performance on the GFMT, and alexithymia score on the TAS. After adjusting for these factors there was still strong evidence for a decrease in emotion recognition after TBI (Model 3, Table 4-3).

An effect size calculation using means and standard deviations for the TBI group compared to the matched control group showed that the effect was large, Hedges' g_s = 1.28. The common language effect size indicated that in this sample the chance of an individual with TBI having decreased emotion recognition compared to an individual from the control group was 82%. Further, there was evidence that better performance on the GFMT was associated with increased performance on the Bristol ERT, but no evidence that any other factors included in the regression models were associated with emotion recognition performance. Collinearity of the variables included in the fully adjusted model (Model 3) was investigated and the variance inflation factor did not exceed 2 for any of the variables, indicating that adjustments for collinearity were not required.

Table 4-3

Regression models for overall emotion recognition on the Bristol ERT

Predictor	*b*	95% CI		*t*	*p*	Model Fit
		LL	UL			
Model 1						n = 52
(Intercept)	0.669	0.634	0.705			R^2 = 0.308
TBI	-0.119	-0.169	-0.068	-4.71	<.001	95% CI [0.11, 0.48]
Model 2						
(Intercept)	0.679	0.638	0.720			n= 52
TBI	-0.112	-0.164	-0.059	-4.27	<.001	R^2 = 0.321
DASS Anxiety score	-0.002	-0.006	0.002	-0.97	0.34	95% CI [0.10, 0.48]
Model 3						
(Intercept)	0.500	0.272	0.724			
TBI	-0.088	-0.140	-0.036	-3.40	0.001	
DASS Anxiety score	-0.003	-0.007	0.001	-1.34	0.19	
Sex (Female)	0.009	-0.043	0.061	0.36	0.72	
Age left education	-0.005	-0.011	0.001	-1.62	0.11	n = 51*
GFMT (total hits)	0.009	0.004	0.014	3.42	0.001	R^2 = 0.485
TAS (total score)	-0.001	-0.002	0.002	-0.13	0.90	95% CI [0.19, 0.58]

Note: TBI is Traumatic Brain Injury. DASS is Depression, Anxiety, and Stress Scale. GFMT is Glasgow Face Matching Test. TAS is Toronto Alexithymia Scale. b is the unstandardised regression coefficient from a multiple linear regression. LL is the lower limit and UL the upper limit of the 95% Confidence Interval. t is the t-test coefficient.
** One patient did not complete the TAS so was dropped at this stage in the analysis.*
Model 1 – Association between TBI and overall emotion recognition measured on the Bristol ERT.
Model 2 - Model 1 adjusted for anxiety using the DASS anxiety subscale score
Model 3 - Model 2 additionally adjusted for sex, age left education, GFMT, and TAS

4.3.3.1 Sensitivity analysis

There were no outliers on the Bristol ERT, but two TBI participants failed the online attention check despite no obvious issues during the testing session. The primary analysis was rerun as a sensitivity analysis with those two participants and their matched controls removed. There was no change in the associations observed using this reduced sample of 24 TBI participants and 24 matched controls (Appendix D, section 8.10).

4.3.4 Secondary analyses

4.3.4.1 Emotion specific outcomes

To investigate emotion recognition accuracy for each individual emotion, the regression analysis was rerun using the unbiased hit rate for each emotion as the outcome. There was evidence that participants with TBI were worse at recognising

anger, disgust, fear, and surprise (Table 4-4). However, when the regression was adjusted for anxiety (Model 2) and then other factors (Model 3), there was no longer any evidence that participants with TBI had decreased emotion recognition accuracy for disgust. Evidence for decreased emotion recognition accuracy for anger, fear, and surprise stayed consistent when the analysis was adjusted for confounders. The same pattern of results was not observed when raw hits were used as the outcome measure instead of the unbiased hit rate (Appendix E, section 8.11). Again, there is evidence that participants with TBI are worse than controls at identifying anger, disgust, and fear, but there is no evidence that they have less hits for surprise and strong evidence for fewer hits for sad. To investigate whether participants with TBI were more/less likely to select a particular emotion on the Bristol ERT compared to matched controls the same regression analysis was conducted using the number of times an emotion was selected as an outcome (Bias Score). There was weak evidence that participants with TBI were less likely to label faces as sad and more likely to label faces as surprised compared to controls (Table 4-4). There was no evidence that anxiety was associated with individual emotion recognition performance on the Bristol ERT (Appendix F, section 8.12).

4.3.4.2 Emotion Bias Tasks

Multiple linear regression was used to investigate whether there was evidence for a difference between the TBI and control groups on the two Emotion Bias Tasks (EBTs) that were completed in addition to the Bristol ERT. There was no evidence for a difference in balance point between participants with TBI and the matched controls on either EBT (angry: b = -0.603, 95% CI [-1.44, 0.238], $t(50) = -1.44$, $p = 0.16$, sad: $b = 0.462$, 95% CI [-0.42, 1.34], $t(50) = 1.05$, $p = 0.3$). When the model was adjusted for anxiety, there was still no evidence for a difference between the TBI and control groups, but there was weak evidence that anxiety was associated with an increased balance point on the sad EBT ($b = 0.07$, 95% CI [0.001, 0.139], $t(49) = 2.04$, $p = 0.05$). However, there was no evidence for this association once the model was additionally adjusted for sex, age that participants left education, performance on the GFMT, and TAS score ($b = 0.05$, 95% CI [-0.027, 0.129], $t(44) = 1.33$, $p = 0.19$). The fully adjusted model also showed that there was weak evidence that better performance on the GFMT was associated with an increased balance point on the angry EBT ($b = 0.099$, 95% CI [0.006, 0.192], $t(44) = 2.15$, $p = 0.03$), and a trend in the same direction for the sad EBT ($b = 0.094$, 95% CI [-0.001, 0.191], $t(44) = 1.99$, $p = 0.05$). There was no evidence for an association between the balance point on either EBT and factors other than GFMT in the full adjusted model.

Table 4-4

Results from regression analysis investigating association between TBI and emotion recognition performance for each emotion

Emotion	Model 1 - Unadjusted					Model 2 - Adjusted for Anxiety					Model 3 - Fully Adjusted				
	b	95% CI		*t*	*p*	*b*	95% CI		*t*	*p*	*b*	95% CI		*t*	*p*
		LL	UL				LL	UL				LL	UL		
UNBIASED HIT RATE (HU)															
Angry	-0.024	-0.036	-0.011	-3.8	< .001	-0.022	-0.035	-0.009	-3.39	< .001	-0.016	-0.03	-0.002	-2.36	0.02
Disgust	-0.014	-0.028	0.000	-2.04	0.05	-0.014	-0.028	0.001	-1.9	0.06	-0.012	-0.027	0.004	-1.48	0.15
Fear	-0.031	-0.047	-0.015	-3.92	< .001	-0.029	-0.046	-0.013	-3.54	< .001	-0.022	-0.037	-0.006	-2.75	0.01
Happy	-0.001	-0.010	0.009	-0.14	0.89	-0.002	-0.012	0.008	-0.38	0.71	-0.005	-0.016	0.005	-1.01	0.32
Sad	-0.008	-0.019	0.003	-1.38	0.17	-0.006	-0.018	0.005	-1.12	0.27	-0.003	-0.016	0.009	-0.51	0.61
Surprise	-0.018	-0.025	-0.010	-4.73	< .001	-0.017	-0.024	-0.009	-4.32	< .001	-0.013	-0.021	-0.005	-3.3	< .001
BIAS SCORE (number of responses given for each emotion)															
Angry	0.31	-1.65	2.27	0.32	0.75	0.48	-1.57	2.53	0.47	0.64	-0.06	-2.34	2.21	-0.06	0.95
Disgust	-1.38	-4.01	1.24	-1.06	0.29	-1.69	-4.43	1.04	-1.24	0.22	-1.18	-4.16	1.8	-0.8	0.43
Fear	0.77	-1.93	3.47	0.57	0.57	0.99	-1.83	3.81	0.71	0.48	1.01	-2.06	4.08	0.66	0.51
Happy	-0.77	-4.42	2.88	-0.42	0.67	-0.15	-3.92	3.62	-0.08	0.94	1.16	-2.73	5.06	0.6	0.55
Sad	-4.12	-7.5	-0.73	-2.44	0.02	-4.14	-7.69	-0.58	-2.34	0.02	-4.43	-8.43	-0.44	-2.24	0.03
Surprise	5.19	2.31	8.07	3.62	< .001	4.5	1.57	7.44	3.09	< .001	3.5	0.42	6.59	2.29	0.03

Note: b is the TBI unstandardised regression coefficient from a multiple linear regression. LL is the lower limit and UL the upper limit of the 95% Confidence Interval. t is the t-test coefficient. In bold are the estimates with evidence for an association of TBI and emotion recognition. Model 1 - Association between TBI and overall emotion recognition measured using the Bristol ERT. Model 2 - Model 1 adjusted for anxiety on the DASS. Model 3 - Model 2 also adjusted for sex, age left education, GFMT, and TAS.

4.3.4.3 Other factors

To investigate whether overall DASS score, DASS depression, DASS stress, or BPAQ score were associated with changes in emotion recognition after TBI, models 2 and 3 of the primary analysis were rerun substituting the DASS anxiety score with each of those factors in turn. There was no evidence for an association of performance on the Bristol ERT after TBI with overall score on the DASS (*b* = -0.0003, 95% CI [-0.001, 0.001], *t*(49) = -0.65, *p* = 0.52), the DASS depression score (*b* = -0. 0001, 95% CI [-0.003, 0.003], *t*(49) = -0.1, *p* = 0.92), DASS stress score (*b* = -0.001, 95% CI [-0.004, 0.002], *t*(49) = -0.8, *p* = 0.43), or the overall score on the BPAQ (*b* = -0.0001, 95% CI [-0.001, 0.001], *t*(49) = -0.27, *p* = 0.79). Further, once fully adjusted (model 3) only TBI and performance on the GFMT showed evidence of an association with performance on the Bristol ERT, as found in the primary analysis.

4.3.5 Reliability

Internal reliability of accuracy on the Bristol ERT was assessed using split-half reliability approach outlined by Parsons, Kruijt and Fox (2019) and by calculating Cronbach's alpha. For the split-half reliability all the trials were treated as the same condition and 5000 random splits were used to establish a Spearman-Brown corrected reliability estimate which corrects for task length. Similarly, the confidence interval for Cronbach's alpha was calculated using 5000 bootstrap samples. Trials for both the 26 participants with TBI and their non-injury matched controls were used to estimate the reliability of the Bristol ERT in this sample. As was outlined in Chapter 2 the Standard Error of Measurement (SEM) and the Smallest Real Difference (SRD) can be helpful to interpret individual scores and will also be calculated. Furthermore, Cronbach's alpha was used to adjust the effect size estimate according to Baugh (2002) using Cohen's d.

In the full sample of 52 participants the Spearman-Brown corrected estimate was r_{sb} = 0.8 (95% CI: 0.71 to 0.87) and Cronbach's alpha for the 96 items was α = 0.86 (95% CI: 0.81 to 0.9). When low intensity trials in the Bristol ERT were removed, the Spearman-Brown corrected reliability estimate was only slightly higher at r_{sb} = 0.82 (95% CI: 0.74 to 0.88) and Cronbach's alpha for the 60 items version was approximately the same at α = 0.86 (95% CI: 0.8 to 0.89). The standard deviation for the full sample on the Bristol ERT was 10.36, which means that the SEM was 4.63 ($10.36\sqrt{1-0.8}$) and the SRD was 12.8 hits (2.77×4.63). The unadjusted Cohen's d for the full sample was 1.32,

which was converted to an r value of 0.55, and using the formulas outlined by Baugh (2002) the corrected Cohen's d was calculated (*d* = 1.35). The adjusted estimate does not change the interpretation, so the unadjusted Hedge's g was used.

Reliability estimates for the TBI and non-injury groups were also calculated separately, so with 26 participants in each group. The Spearman-Brown corrected reliability estimate was slightly lower in the TBI group compared to the no injury group, but there was considerable overlap in the confidence intervals (r_{sb} = 0.7, 95% CI: 0.5 to 0.84, and r_{sb} = 0.74, 95% CI: 0.57 to 0.86, respectively). The SEM for the TBI group in this study was 4.71 ($8.6\sqrt{1-0.7}$) and in turn the SRD was 13 hits (2.77×4.71). The SEM for the control group was 4.49 ($8.8\sqrt{1-0.74}$) and in turn the SRD was 12.4 hits (2.77×4.49). This means that participants would have to show a change in performance of 13 hits to be confident of a change in performance.

4.4 Discussion

The main aim of the current study was to evaluate whether the Bristol ERT can be effectively used to assess emotion recognition in a moderate to severe TBI population. Our main hypothesis was that the Bristol ERT could be used to replicate findings from a feasibility study (Müller-Glodde, 2015) and prior research (Babbage et al., 2011) showing that people with TBI perform worse than neurologically healthy controls on overall emotion recognition. There was strong evidence for decreased overall emotion recognition on the Bristol ERT in the TBI group even after controlling for potentially confounding factors. Contrary to our prediction, participants with TBI did not show decreased emotion recognition accuracy for all six emotions, only for anger, fear, and surprise. The secondary aim of this study was to develop our understanding of the associations between performance on the Bristol ERT and potentially confounding factors in a TBI population, in particular considering anxiety as a possible moderator. There was evidence that levels of anxiety were higher in the TBI group compared to the control group, however no evidence that anxiety was associated with emotion recognition accuracy or emotion recognition bias in this sample once confounding factors had been controlled for.

4.4.1 Moderate to severe TBI and Emotion Recognition

In line with the current literature (Babbage et al., 2011; Murphy et al., 2021), results from this study indicated that having moderate to severe TBI was associated with decreased emotion recognition. The size of the effect observed in this study was comparable to the one calculated by Babbage and colleagues (2011) in their meta-analysis, which indicates that the Bristol ERT is suitable for use in a moderate to severe TBI population. The results also suggested that the lower emotion recognition scores in the TBI group were not explained by the potentially confounding factors of anxiety, face perception, alexithymia, depression, stress, or aggression in this sample.

4.4.1.1 Individual emotion recognition accuracy

The two emotions most consistently reported as being associated with lower scores in TBI samples compared to controls are anger and disgust (Rigon et al., 2016; Rosenberg et al., 2015; Rosenberg et al., 2014). There was evidence in this study that participants with TBI were less accurate at identifying angry faces, but not disgusted faces after controlling for confounding factors. Instead, there was evidence that

participants with TBI in this study were also less accurate at identifying fearful and surprised faces. It is possible that the difference in findings is due to the use of an unbiased hit rate on the Bristol ERT instead of simply using the number of correct responses made for each emotion. Callahan and colleagues (2011) have suggested that people who have experienced a TBI could be more likely to attribute negative emotions to ambiguous stimuli. This would impact the misidentifications made and in turn influence the emotion recognition accuracy measured using the unbiased hit rate. Arguably fearful faces are commonly misidentified as surprised (Rosenberg et al., 2014). This could explain why the bias scores showed that participants with TBI in this sample were more likely to label faces as surprised than control participants. It is likely that the low emotion recognition accuracy based on the unbiased hit rate for both fearful faces and surprised faces associated with the TBI group was the result of patients labelling many fearful stimuli as surprised. One participant with TBI mentioned at the end of their session that they labelled a face as surprised if they were confused about the emotion presented. Further research is needed to explore whether surprise is more commonly selected by participants with TBI if stimuli are ambiguous or hard to read. The response bias scores in this study also indicated that participants with TBI were less likely to label emotions as sad, but interestingly there was no evidence that they were less accurate than controls at identifying sad faces based on the unbiased hit rate. However, when raw hits are considered, there was strong evidence that participants with TBI had a lower number of correct responses for sad faces compared to the control group. These results highlight the issue that on six alternate force tasks, such as the Bristol ERT, there is only one target versus five distractors for any given trial. As is obvious for the emotion of sadness in this sample, this means that correct responses and misidentifications for any given emotion do not necessarily mirror one another (Megreya & Burton, 2007). The results from this study indicate that unbiased hit rate may not be a good alternative to using hits as a measure of individual emotion recognition accuracy. The implications of this will be discussed in further detail in the general discussion of this thesis.

4.4.1.2 Emotion bias

There was no evidence that having a moderate to severe TBI was associated with bias on either of the two Bristol EBT suggesting that participants with TBI did not show a negative attribution bias in this sample. Neumann, Sander and colleagues (2021) have shown that a negative attribution bias towards angry faces is associated with

increased levels of aggression in a TBI population. There was no evidence that participants with TBI in this sample had increased levels of aggression based on the BPAQ, which could explain the lack of association observed. One of the hypotheses made for this study was that anxiety would be associated with negative attribution bias on both Bristol EBT tasks. It is interesting that there was no evidence for a difference in performance on the Bristol EBT tasks despite evidence for increased levels of anxiety in the TBI group compared to the controls. See the next section for further discussion.

4.4.2 Associations with Covariates

4.4.2.1 Anxiety

The trend towards increased anxiety measures using the DASS in the TBI group is consistent with current literature about increased clinical anxiety without diagnosis of anxiety disorders after TBI (Osborn et al., 2016). However, there was no evidence that anxiety was associated with emotion recognition in this sample after having a TBI was controlled for. This suggests that increased prevalence of anxiety after TBI does not explain the deficits in emotion recognition observed in this population. This finding does not match the prediction that anxiety would be associated with a decrease in emotion recognition accuracy and an increased negative attribution bias. Although having an anxiety disorder has been associated with decreased emotion recognition accuracy these effects have been small unless participants were presented with symptoms of PTSD (Plana et al., 2014). It is possible that there was no association between anxiety and emotion recognition and anxiety in this study because the DASS is a more general measure of clinical anxiety. Items on the DASS are rated based on symptoms of anxiety experienced within the last week and does not capture whether participants are in an acute state of anxiety. State anxiety has been associated with both decreased emotion recognition and negative attribution bias on the Bristol ERT and EBT (Attwood et al., 2017; Dyer et al., 2022). Dyer and colleagues (2022) reported that only participants with hight trait anxiety presented with a tendency to identify ambiguous stimuli as angry on the Bristol EBT when state anxiety was induced. Given the evidence of increased levels of anxiety after TBI it is possible that emotion recognition would be impaired when they are in a state of acute anxiety, which in turn could affect their social interactions.

4.4.2.2 Face perception and Alexithymia

There was clear evidence of an association between emotion recognition performance on the Bristol ERT and ability to discriminate between faces as measured using the GFMT. This was expected given that the Bristol ERT is a facial emotion recognition task and even the Emotion Evaluation Test from TASIT, which does not solely rely on emotion recognition from faces, has been associated with performance on face perception tasks (McDonald et al., 2006). The GFMT was included in this study to help evaluate whether differences in emotion recognition between the TBI and control groups could be explained by a general difficulty in identifying faces (J. D. Henry et al., 2015). There was no evidence that the TBI and control groups differed in performance on the GFMT, and results clearly indicated that the lower emotion recognition scores associated with the TBI were not explained by participants' ability to perceive faces. Although the GFMT is not a test of general cognitive ability, this finding matches findings that cognitive ability associated with but does not fully explained performance on emotion recognition task (Rosenberg et al., 2015). A slightly unexpected finding was that there was a trend towards higher GFMT scores being associated with more positive responses on both Bristol EBT tasks. All the stimuli included on the Bristol EBT are ambiguous, so there is no right or wrong answer on any given trial. The fact that participants with better performance on the GFMT were more likely to identify faces as happy is difficult to explain. It is possible that this was a spurious association, as the analysis was exploratory and evidence for the association was weak.

There was weak evidence that having a TBI was associated higher alexithymia scores which is consistent with prior research indicating increased alexithymia after TBI (Henry et al., 2006; Williams & Wood, 2010). It has been suggested that increased levels of alexithymia could account for difficulties in emotion recognition, especially when ambiguous stimuli are only presented for a short period of time, as during the Bristol ERT (Ihme, Sacher, Lichev, Rosenberg, Kugel, Rufer, Grabe, Pampel, Lepsien, Kersting, Villringer, Lane, et al., 2014; Wearne et al., 2019; Williams & Wood, 2010). Despite the increased levels of alexithymia in the TBI group compared to controls, there was no evidence that alexithymia was associated with performance on the Bristol ERT in this study. This is consistent with findings that scores on the TAS-20 are not correlated with performance on the Emotion Evaluation Test from TASIT (McDonald et al., 2011; Rosenberg et al., 2019). This suggests that alexithymia may not be linked to

performance on emotion labelling tasks, which raises the question of what constructs the emotion labelling tasks are tapping into.

4.4.3 Reliability of the Bristol ERT

Internal consistency reliability estimates for the Bristol ERT in this study were high based on the descriptors used by Strauss and colleagues (2006). Both the split half reliability estimates and Cronbach's alpha were above 0.8, which has been suggested as a minimally acceptable threshold of reliability for tasks used in individual assessment (Ponterotto & Ruckdeschel, 2007; Sattler, 2001). This shows that the measurements taken using the Bristol ERT in this sample are likely to result in reliable inferences and indicates that reliability may be sufficient for use in a clinical setting. Reliability estimates for the TBI group and control groups separately were lower, which is not surprising given the reduced sample sizes are likely to result in more noise in the data. The fact that the reliability estimates across the two groups were comparable suggests that it is appropriate to use the Bristol ERT in a TBI population. Nevertheless, it is important to note that the Smallest Real Difference estimate indicated that a big change (± 13 hits) in individual performance is needed to be confident in a true change. Wider implications and potential task adjustments to increase reliability estimates for the Bristol ERT are presented in the general discussion (Chapter 7).

4.4.4 Strengths and Limitations

This is the first study using the Bristol ERT to assess emotion recognition in a TBI population. It clearly shows that the Bristol ERT is suitable for research in this population and that it is worth considering whether the task is useful for individual assessment of emotion recognition after TBI. Furthermore, several confounding factors were included in the analysis and there was no evidence that any of these explain the association between emotion recognition and moderate to severe TBI. However, a clear limitation of this study was that due to insufficient data it was not possible to investigate the association between general cognitive ability and performance on the Bristol ERT in a meaningful way. The short presentation time for stimuli during the Bristol ERT could make the task vulnerable to deficits in processing speed and attention. The neuropsychological test scores available for patients with TBI showed that processing speed, attention, and memory scores were below population average suggesting that general cognitive ability could have contributed to the decreased emotion recognition scores. Future studies should investigate whether processing speed and attention could

underlie the association between performance on the Bristol ERT and having a TBI. The fact that all the participants were recruited from a rehabilitation setting means that the sample could be biased towards people with worse outcomes after moderate to severe TBI. This means general cognitive ability could be particularly critical in this sample.

It is important to consider that this study was designed to detect the main effect of decreased overall emotion recognition performance after TBI, not emotion specific outcomes. Notably, if a Bonferroni correction is applied to the emotion specific analysis, there was no longer any evidence for emotion specific effects in this study aside from decreased accuracy for surprised faces. Using this method of correction has been criticised as being overly conservative in small samples (VanderWeele & Mathur, 2019) so has not been applied in this thesis. Instead of using the p-value as a dichotomous indicator the results are discussed in terms of the strength of evidence against the null hypothesis (Colling & Szucs, 2021; Lakens et al., 2018). Also, the study design does not allow for causal inferences to me made as there was no way to assess emotion recognition prior to the TBI. Investigating the causal pathway between emotion recognition and TBI is difficult and would require a long-term prospective study where measures of emotion recognition are available pre and post injury. Likely due to the cost and timeframe for such a study this type of research has not yet been conducted but would be useful to consider in the future.

4.4.5 Implications and future directions

These results indicated that the Bristol ERT is sensitive to changes in emotion recognition associated with moderate to severe TBI, which suggests that the task could be useful as part of neuropsychological assessments. However, further work is needed to develop the Bristol ERT as a measure of individual performance. The results also showed that whilst total hits was a suitable measure of overall emotion recognition, the individual emotion outcome measures require further development. Potentially using discriminability thresholds would be a better approach to use (Cecilione et al., 2017). A possible next step would be to identify a group of participants with TBI that are reported to struggle with social interactions and compare their performance on the Bristol ERT to a matched group of participants with TBI who are not reported to experience social communication or interaction difficulties. This would help determine a minimal important difference in performance (Anvari & Lakens, 2021).

Chapter 5 Assessing emotion recognition after mild Traumatic Brain Injury using the Bristol Emotion Recognition Task

5.1 Introduction

Moderate to severe TBI has been clearly linked to emotion recognition deficits (Babbage et al., 2011; Murphy et al., 2021) but few studies to date have investigated the association between mild TBI and emotion recognition (Theadom et al., 2019). The study presented in this chapter used the same study design as the study in Chapter 4 to investigate the association between mild TBI and emotion recognition on the Bristol Emotion Recognition Task (ERT). This will allow for a direct comparison of the two studies and help further our understanding of emotion recognition in a mild TBI population. The study was registered on the Open Science Framework and sections are taken from the pre-registered protocol (10.17605/OSF.IO/29XA4).

5.1.1 Mild TBI and Emotion Recognition

Indicators of injury severity injury seem to be negatively correlated with performance on emotion recognition tasks suggesting people with less severe TBI injuries tend to perform better on emotion recognition tasks (Ietswaart et al., 2008; Spikman et al., 2012). This could explain why most studies to date have investigated the association between emotion recognition and moderate to severe TBI but not mild TBI (Calvillo & Irimia, 2020). Theadom and colleagues (2019) have recently addressed this gap in the literature by using The Awareness of Social Inferences Test (TASIT) to assess social cognition in adults four years after mild TBI. They found that participants with mild TBI scored lower than neurologically healthy controls on overall social cognition but did not find evidence for a group difference in overall emotion recognition performance on the Emotion Evaluation Test (EET) subscale from TASIT. Analysis of performance on each individual emotion indicated that participants with mild TBI were less accurate at identifying happy stimuli compared to controls. This suggests that the task is sensitive to changes in emotion recognition after mild TBI, and that unlike moderate to severe TBI, mild TBI may be associated with emotion specific changes but not deficits in overall emotion recognition.

5.1.2 Current study: Rationale

This current study investigated associations between emotion recognition on the Bristol ERT and mild TBI in a community-based sample. Use of morphed stimuli at

different intensities could help detect subtle differences in emotion recognition and will help develop our understanding of potential associations between mild TBI and emotion recognition. Participants were screened for TBI before being invited to complete the full study. If participants met the criteria for mild TBI defined as a head injury with loss of consciousness, they were included in the TBI group, whilst participants without a history of TBI or other neurological disorders were included as a control group. This approach was used to maximise specificity, i.e., increase likelihood that participants in the TBI group were correctly identified as having had a TBI. Additionally, information about severity of the TBI was collected and considered as part of the analysis.

As discussed in the introduction to this thesis there are many factors that could influence emotion recognition after TBI and could confound the association between mild TBI and emotion recognition. This study considered anxiety (Attwood et al., 2017; Plana et al., 2014), age (Byom et al., 2019), sex (Rigon et al., 2016; Zupan et al., 2017), face perception (J. D. Henry et al., 2015), and alexithymia (Wearne et al., 2019) as potential confounders for the association between emotion recognition and TBI. Depression (Alway et al., 2016; Dalili et al., 2015; Watters & Williams, 2011) and aggression (Hoaken et al., 2007; Neumann et al., 2017) have also been associated with changes in emotion recognition and are prevalent after TBI. Although not the focus of this study, participants also completed questionnaires measuring those two factors. Of particular interest was the potential impact of anxiety on the association between emotion recognition and TBI.

Anxiety has been associated with deficits in emotion recognition accuracy (Attwood et al., 2017; Plana et al., 2014), which could be contributing to deficits in emotion recognition associated with TBI. Theadom and colleagues (2019) assessed whether anxiety was associated with overall performance on TASIT after mild TBI and did not find evidence for an association. Notably, they used a measure of anxiety collected at one month post injury not at the time that social cognition was assessed four years later. It is possible that they did not find an association because levels of anxiety change over time and are most prevalent in the first year (Ponsford et al., 2018). Given that state anxiety has been associated with deficits in emotion recognition (Attwood et al., 2017; Dyer et al., 2022) a measure of anxiety should be included at time of testing to assess the associations between mild TBI, emotion recognition, and anxiety. Anxiety has also been associated with negative attribution bias, meaning that ambiguous emotional stimuli are interpreted as being more negative or threatening

(Attwood et al., 2017; Mendes Ferrer Rosa et al., 2017). This means that assessing emotion recognition accuracy and bias in conjunction could provide useful insights about changes in emotion recognition associated with mild TBI. An advantage of using the Bristol ERT compared to the EET from TASIT used by Theadom and colleagues (2019) to investigate emotion recognition after mild TBI is that both emotion recognition accuracy and response bias can be assessed. Furthermore, an Emotion Bias Task (EBT) based on the same stimuli as the Bristol ERT (See section 4.2.3.4 under method for details) was developed as a more sensitive measure of emotion bias and was also included in this study.

5.1.3 Current study: Aims and Hypotheses

The main aim of the study was to establish whether there is evidence for an association between mild TBI and emotion recognition. Our primary hypothesis was that mild TBI would be associated with a deficit in overall emotion recognition based on the evidence for a deficit in emotion recognition after moderate to severe TBI.

H1 Participants who report having had a TBI will be less accurate at recognising emotions on the Bristol ERT than neurologically healthy controls.

Additionally, this study was designed to assess the potential impact of confounding factors that are commonly reported after TBI, with a focus on the potential impact of anxiety. Anxiety was predicted to be associated with deficits in emotion recognition and negative attribution bias. Regarding emotion specific recognition accuracy, the hypothesis was that deficits in emotion recognition accuracy associated with mild TBI would not be emotion specific. The secondary hypotheses made in the pre-registered protocol were:

H2: Increased anxiety will be associated with a decreased emotion recognition accuracy on the Bristol ERT.

H3: Increased anxiety will be associated with a negative attribution bias towards anger and fear on the Bristol ERT.

H4: The difference in emotion recognition accuracy between participants with TBI and neurologically healthy controls persists after adjusting for age, sex, face perception and emotion processing.

H5: The difference in emotion recognition accuracy between participants with TBI and neurologically healthy control participants is not emotion specific.

Inclusion of the Bristol EBT in this study was considered exploratory and no specific hypotheses were pre-registered.

5.2 Method

5.2.1 Design

The study had an observational, cross-sectional design to investigate the associations between TBI, anxiety, and emotion recognition. Performance on a six-alternative forced-choice emotion recognition task (Bristol ERT) was compared between a group of participants who reported having a TBI and a group of participants without TBI. Overall emotion recognition accuracy (total hits) was used as the primary outcome, with emotion specific recognition accuracy, and emotion specific responses as secondary outcomes. A measure of bias in detecting 'sad' and 'angry' faces compared to 'happy' faces was also used as a secondary outcome. Self-report measures of anxiety, mood, face discrimination, alexithymia, and aggression were completed at the time of testing. There was no element of randomisation in the study design. Ethics approval was obtained from the School of Psychological Science Research Ethics Committee at the University of Bristol (011020111402). The study was preregistered on the Open Science Framework (10.17605/OSF.IO/29XA4).

5.2.2 Participants

Participants registered on Prolific (https://www.prolific.co/participants) were able to see the study if they matched the study criteria. The screening criteria for all participants included age specifications, confirmation of visual ability, UK residency, and English language fluency (See Appendix A, section 8.13).

5.2.2.1 Participants with Traumatic Brain injury

Participants with TBI were recruited from a pool of Prolific participants who had indicated that they have had a head injury but did not report other neurological conditions or chronic illness. A brain injury screening questionnaire was used to establish that all participants with TBI included had a head injury with loss of consciousness (See section 5.2.2.3).

Inclusion criteria for participants with TBI

- Report having had a TBI with loss of consciousness
- Above the age of 18 years
- Normal or corrected to normal vision
- Resident in the UK and fluent in English

Exclusion criteria for participants with TBI

- Report a history of neurological conditions other than TBI (including but not limited to stroke, MS, epilepsy, autism, or brain tumors)
- Report a history of schizophrenia or psychosis
- Currently under the influence of substances that can substantially alter perception (e.g. psychoactive drugs or excessive alcohol)

5.2.2.2 Neurologically healthy control participants

Neurologically healthy controls were recruited from Prolific participants who did not report having had a head injury, neurological conditions, or chronic illness.

Inclusion criteria for neurologically healthy control participants

- Above the age of 18 years
- Normal or corrected to normal vision
- Resident in the UK and fluent in English

Exclusion criteria for neurologically healthy control participants

- Report a history of neurological conditions (including but not limited to TBI, stroke, MS, epilepsy, autism, or brain tumors)
- Report a history of schizophrenia or psychosis
- Currently under the influence of substances that can substantially alter perception (e.g. psychoactive drugs or excessive alcohol)

5.2.2.3 Brain Injury Screening

Participants who were eligible to sign up to the study based on the Prolific screening questions were given access to an information sheet and invited to complete a brain injury screening questionnaire (See measures section 5.2.3.1 for details). They were required to complete a consent form before completing the screening questions and asked at the end whether they still wanted to be considered for the full study. Participants who confirmed not having had a head injury or other neurological conditions were eligible to complete the full study as neurologically healthy controls. Participants who reported having had a head injury and received a score of 5 or more on the Brain Injury Screening Index were eligible to complete the full study as participants with TBI. All participants who completed the brain injury screening were reimbursed

£0.50 for their time regardless of eligibility to take part in the full study. Eligible participants were invited to sign up to the full study and received an additional £6.50 once completed.

5.2.2.4 Sample size determination

A pilot study investigating performance on the Bristol ERT after brain injury found an effect size of d = 1.22 for the difference in emotion recognition accuracy between participants with brain injury and non-brain injury controls (Müller-Glodde, 2015). This was comparable to the effect size reported for the difference in emotion recognition after TBI reported in a meta-analysis by Babbage and colleagues (2011). As effect sizes from small sample studies can be inflated (Button et al., 2013), the observed effect size was reduced by one third for the purpose of the sample size calculation. A sample size calculation was conducted using the *pwr* package (Champely, 2020) for R version 4.0.2 (2020) in RStudio (2020). Seven predictors were included in the regression analysis, namely, self-reported TBI, anxiety, sex, age, years of education, face discrimination, and emotion labelling. A significance level of 0.05, and 80% power were considered adequate, and the reduced Cohen's d of 0.8 was converted into an f^2 value of 0.16. Based on these parameters 97 participants were required for this study. Consequently, the minimum sample size for this study was set at 100, with 50 participants who reported having had a TBI and 50 neurologically healthy controls.

5.2.3 Measures

The tasks and questionnaires for this study were coded in and presented through Gorilla Experiment Builder (https://gorilla.sc/; Anwyl-Irvine et al., 2019). Measures are listed in the order in which they were presented to participants. Aside from the brain injury questionnaire and demographics questionnaire all the measures included were also used in the moderate to severe TBI study discussed in Chapter 4. Please refer to the methods of that chapter (Section 4.2.3) for details on the tasks.

5.2.3.1 Brain Injury Screening Index (BISI)

The Brain Injury Screening Index (The Disabilities Trust, 2018) available at https://www.thedtgroup.org/foundation/brain-injury-screening-index as adapted for online presentation in this study. The BISI is a validated self-report measure of brain injury (Ramos et al., 2020). Participants were asked whether they have had a serious blow to the head, and or other neurological illnesses. If they had experienced a TBI

(defined as a serious blow to the head) they were asked about the nature of the injury, time since injury, whether they experienced loss of consciousness, and whether they experienced a period of amnesia around the incident of injury. The primary outcome was the sum of scores assigned to questions 1, 3, 4, 5, 7, 8, and 10 on the questionnaire. To maximise specificity a score of 5 or more on the questionnaire is recommend as positive screen for brain injury, which was used as the cut-off in this study. However, any participants who reported neurological illness other than TBI (question 8) were not eligible to complete the full study. Participants who reported having had a serious blow to the head were additionally asked whether they had been given a diagnosis of TBI by a healthcare professional.

5.2.3.2 Demographics questionnaire

Participants were asked to provide information about age, gender, years of education, occupation, and given the option to declare reasons why their emotion recognition might be affected. An attention check was also included asking participants to respond yes to show they had read that question.

5.2.3.3 Screening task

Please refer to Chapter 4 section 4.2.3.2.

5.2.3.4 Bristol Emotion Recognition Task (ERT)

The Bristol ERT is described in detail in the introductory chapter (Section 1.3.2). For details of the task used in the current study please refer to Chapter 4 section 4.2.3.3.

5.2.3.5 The Bristol Emotion Bias Task (EBT) – Happy - Anger & Happy - Sadness

Please refer to Chapter 4 section 4.2.3.4.

5.2.3.6 Glasgow Face Matching Test – short version (GFMT)

Please refer to Chapter 4 section 4.2.3.5.

5.2.3.7 Depression, Anxiety, and Stress Scale - 42 (DASS)

Please refer to Chapter 4 section 4.2.3.6.

5.2.3.8 Toronto Alexithymia Scale - 20 (TAS)

Please refer to Chapter 4 section 4.2.3.7.

5.2.3.9 Buss-Perry Agression Questionnaire (BPAQ)

Please refer to Chapter 4 section 4.2.3.8.

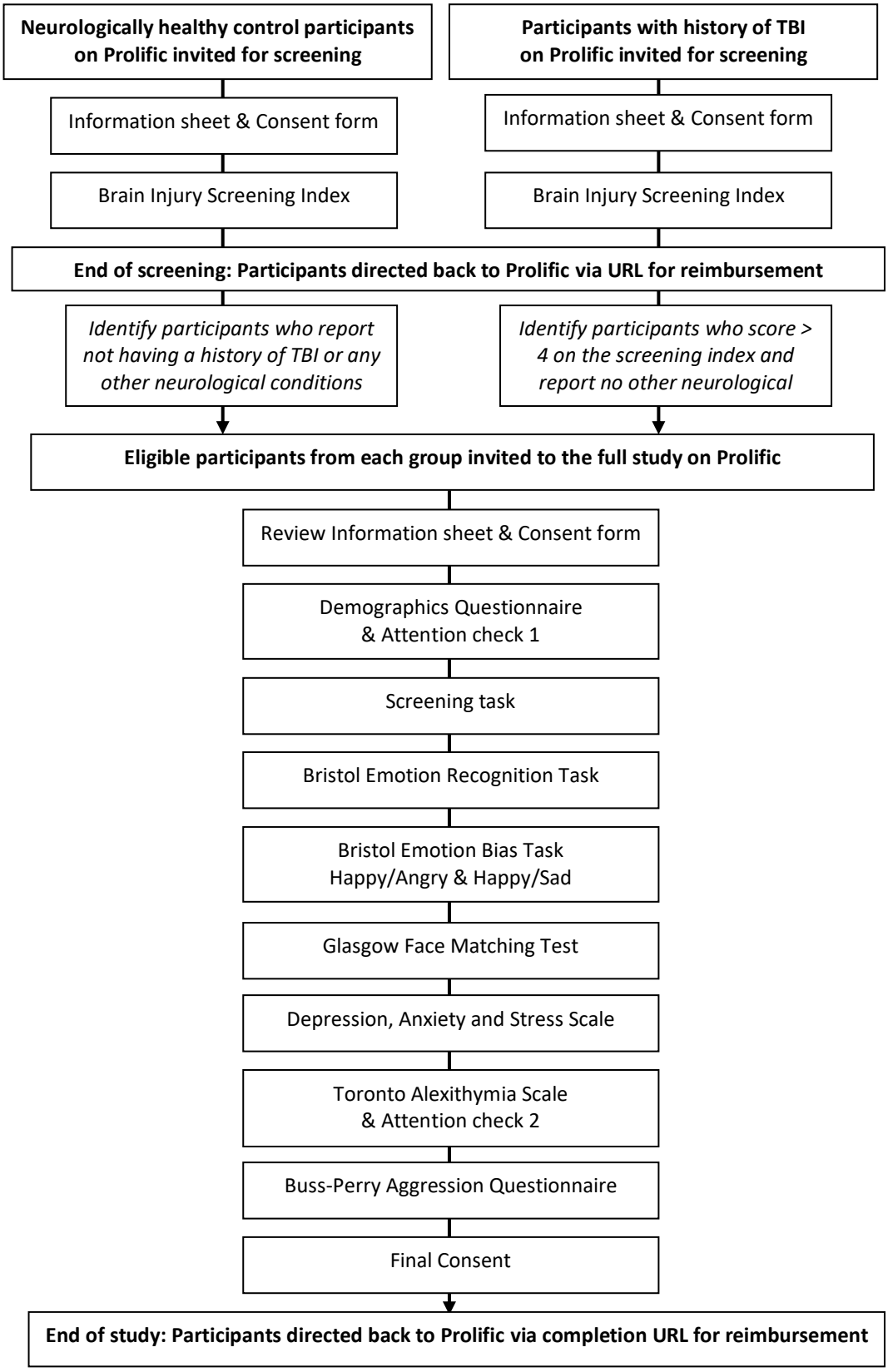
5.2.4 Procedure

Participants who fit the screening criteria set on Prolific were given access to the information sheet and could contact the study team through Prolific if they had any questions. The only time constraint on signing-up to the study by completing the brain injury screening was that data collection ended once the target sample size had been reached. Participants who chose to take part in the study were directed to Gorilla Experiment Builder to complete a consent form and the Brain Injury Screening Index. At the end of the screening, participants were given a completion URL used to authorise reimbursement for the screening through Prolific. The information provided during the screening was used to identify eligible participants, so recruitment was conducted in steps until sufficient participants matching the eligibility criteria had been recruited. See Figure 5-1 for an overview of the study procedure.

All eligible participants were invited to take part in the full study through a message on Prolific and given access to the study using a custom allow list. Before completing the tasks and questionnaires on Gorilla Experiment Builder participants were asked to reconfirm their consent to take part in the study. Tasks and questionnaires were completed in the order described in the study flow chart (Figure 5-1). Participants were given a maximum of 132 minutes to complete the study, which was estimated to take less than an hour, so they had opportunities to take short breaks should they require them. At the end of the study, participants were asked to give consent for the information provided to be used as outlined in the information sheet. A completion URL directed participants back to Prolific and authorised reimbursement for the study.

Figure 5-1

Study Flow Chart



5.2.5 Data Analysis

The data was loaded into RStudio (2020) R version 4.0.4 for data cleaning and analysis. The main analysis packages used were the core *stats* package in R, *psych* package version 2.1.9, *apaTables* package version 2.0.5 (Stanley, 2018), *olsrr* package version 0.5.3 (Hebbali, 2017), *splithalf* package version 0.7.2 (Parsons, 2020), and *ltm* package version 1.1-1 (Rizopoulos, 2006).

5.2.5.1 Data checks

Participants who scored less than 80% of trials correct on the screening task were excluded from the analysis, as they were unlikely to be able to complete the Bristol ERT if they cannot accurately complete the vision screen. Outliers on the Bristol ERT were determined using a box plot, to identify points more than 1.5 times the Inter Quartile Range below the 1st quartile or above the 3rd quartile. A sensitivity analysis was conducted excluding participants who were considered outliers on the Bristol ERT, and or failed one or both attention checks. This means the primary analysis re-run without potential outliers and possibly low-quality data to check for qualitative differences.

5.2.5.2 Primary analysis

Linear regression was used to investigate the association between TBI and overall performance accuracy on the Bristol ERT (total hit rate). The unadjusted model included only TBI as a predictor of emotion recognition, where TBI was coded as a binary variable (participants with moderate to severe TBI and neurologically healthy controls). The second model was adjusted for anxiety using the score for the anxiety subscale on the DASS. Then, the model was additionally adjusted using the demographic data collected, by including participants' age, sex, and years of education as variables in the regression. Finally, the model was adjusted for face perception and alexithymia. The final model included the following variables, TBI, DASS anxiety score, age, sex, years of education, GFMT total score, and TAS score.

5.2.5.3 Secondary analyses

To check for emotion specific changes in emotion recognition after TBI the primary regression analysis outlined above was repeated using emotion specific performance accuracy (unbiased hit rate) and response bias (bias score) on the Bristol

ERT as outcomes. The aim was to check for potential emotion recognition effects that are hidden when only considering overall performance. To investigate how various factors may impact bias in facial affect recognition, the primary regression analysis was re-run using the balance points for 'angry' and 'sad' faces calculated from the Bristol EBT responses. Specifically, the impact of anxiety on the subjective perception of 'sad' and 'angry' expressions compared to 'happy' faces was of interest. Multiple linear regression analyses were also used to explore the potential influence of depression, stress, and aggression on emotion recognition in this sample.

5.2.5.4 Other analyses – deviations from the pre-registered protocol

The pre-registered protocol for this study stated that a series of regressions would be used to investigate a mediating effect of anxiety on the association between TBI and emotion recognition. The mediation analysis was not conducted as there was no evidence for an association between emotion recognition and anxiety in this sample. Further, instead of using false alarms as a measure of response bias, the total number of responses made for each emotion on the Bristol ERT was considered more appropriate. Three analyses were conducted that were not included in the pre-registered protocol. A sensitivity analysis for the primary analysis was conducted excluding participants that were not clearly identifiable as having mild TBI based on loss of consciousness. Further, the TBI group was split into two groups, those who reported Post Traumatic Amnesia and those who did not. The primary analysis was rerun comparing those two groups against performance of neurologically healthy controls. Finally, the reliability of performance on the Bristol ERT was investigated using split half reliability and Cronbach's alpha. Reliability for the whole sample and reliability for the participants with TBI and the neurologically healthy participants separately was assessed.

5.3 Results

5.3.1 Brain Injury Screening

A total of 135 participants completed the Brain Injury Screening, 75 participants with head injury and 60 neurologically healthy controls. 62 of the 75 participants in the head injury group and 57 of the 60 participants in the control group were eligible for the full study. Two participants in each group chose to withdraw from the full study. Five of the participants with head injury were excluded from the analysis after completing the study because they reported loss of consciousness as the reason for hitting their head, so it was unclear whether the head injury resulted in loss of consciousness. The final data set consisted of 110 participants, 55 participants with a self-reported TBI and 55 neurologically healthy controls.

5.3.2 Descriptive statistics

Of the 110 participants, 47 were male and 63 female, however these were not equally distributed between the TBI (37 males) and the control group (10 males). A chi squared test confirmed that there is evidence for a difference in sex across the TBI and control groups ($\chi^2 = 25.1$, $p < 0.001$). Mean age for the TBI group was 37.6 years (12.9) and 35 years (13.8) for the control group and there was no evidence of a difference in age between the groups ($F(1, 108) = 1.02$, $p = 0.32$). Participants in the TBI group on average left education at 21.5 years of age (5.8) and participants in the control group left at around 21.2 years of age (3.2). The mean age that participants left education was comparable across the groups ($F(1, 108) = 0.15$, $p = 0.7$). Descriptive statistics and comparison in performance across the two groups on all the questionnaires and tasks completed as part of this study are presented in Table 5-1.

There was evidence that participants with TBI had worse outcomes on all the questionnaires. They had higher overall DASS scores, and in particular their anxiety and stress scores were higher than those of the control group. Participants with TBI also scored higher on the TAS. Subsequent analysis of the subscales showed that there was strong evidence for a difference on Externally Oriented Thinking ($F(1, 108) = 12.5$, $p < 0.001$), and a trend towards worse performance in the TBI group on the other two subscales; Identifying Feelings ($F(1, 108) = 3.63$, $p = 0.06$), and Describing Feelings ($F(1, 108) = 3.05$, $p = 0.08$). There was weak evidence that the TBI group had a higher aggression score on the BPAQ compared to the controls.

There was no evidence that the groups differed in their performance on any of the cognitive tasks, aside from the Bristol ERT. The data was investigated using a correlation matrix including demographic factors and main outcomes for all the tasks and questionnaires completed (Appendix B, section 8.14). There was evidence that performance on the Bristol ERT was negatively correlated with age and positively correlated with performance on the GFMT. Further, there was evidence for a negative correlation between age and score on the DASS anxiety subscale, and performance on the GFMT was negatively correlated with score on the TAS. There was also evidence that the questionnaire outcomes were all positively correlated with each other. Notably, there was no evidence of an association between the DASS anxiety subscale and overall emotion recognition on the Bristol ERT.

Table 5-1

Performance on questionnaires and cognitive tasks

Variable	TBI				Controls				Group comparison	
	n = 55				n = 55					
Questionnaires	**Mean**	**SD**	**Min**	**Max**	**Mean**	**SD**	**Min**	**Max**	***F-value***	***p***
DASS Total	38.7	29	0	101	26.5	21	0	81	6.40	0.01
Depression	13.1	11	0	42	9.64	9.6	0	35	3.00	0.09
Anxiety	9.53	8.4	0	28	5.76	5.8	0	22	7.48	0.01
Stress	16.0	11	0	38	11.1	8.9	0	29	6.90	0.01
TAS Total	51.0	12	24	77	44.2	12	24	71	8.40	0.01
BPAQ Total	69.6	20	37	107	61.7	18	32	126	4.82	0.03
Cognitive tasks	**Mean**	**SD**	**Min**	**Max**	**Mean**	**SD**	**Min**	**Max**	***F-value***	***p***
Bristol ERT *(total hits)*	62.6	8.0	48	79	65.5	7.9	44	78	3.78	0.05
EBT Angry *(balance point)*	7.01	1.4	4.7	11.3	6.92	1.1	4.7	9.3	0.00	1.00
EBT Sad *(balance point)*	6.97	1.3	4.3	11.3	6.97	1.2	3.3	10	0.12	0.73
GFMT *(total hits)*	31.2	5.4	18	40	32.7	4.5	23	40	2.58	0.11

Note: TBI is Traumatic Brain Injury. DASS is Depression, Anxiety, and Stress Scale. TAS is Toronto Alexithymia Scale. BPAQ is Buss-Perry Aggression Questionnaire. ERT is Emotion recognition Task. EBT is Emotion Bias Task. GFMT is Glasgow Face Matching Test. SD is the standard deviation. Group comparisons conducted using analysis of variance

5.3.3 Traumatic Brain Injury Information

All the neurologically healthy controls scored zero on the Brain Injury Screening. All TBI participants scored six or above on the Brain Injury Screening Index, range 6 to

12. They all reported having lost consciousness and having felt dizzy, unsteady, or dazed at the time of injury. 26 out of 55 participants indicated that they did not remember what had happened in the hours after the injury, which was classified as having had Post Traumatic Amnesia (PTA). Mean time since injury was 201 months (SD = 160) with the most recent injury being 10 months ago and the longest time since injury 670 months. The TBI sample as a whole was classified as mild TBI. Only two participants reported loss of consciousness longer than 30 minutes (but no longer than 1 hour) and 6 participants were unsure about length of loss of consciousness. Only one of the participants with loss of consciousness longer than 30 minutes reported having a diagnosis of TBI (no indication of severity). A sensitivity analysis of the primary regression analysis was conducted with the eight participants who had or may have had loss of consciousness longer than 30 minutes removed (Appendix C, section 8.15).

5.3.4 Primary analysis

A multiple linear regression with overall emotion recognition on the Bristol ERT as the outcome showed that there was weak evidence for a difference in performance between the TBI and control groups (Model 1, Table 5-2). The evidence for decreased emotion recognition after TBI was slightly stronger once the model has been adjusted for anxiety (Model 2, Table 5-2), but there was no evidence for this association once the model was additionally adjusted for demographic factors, performance on the GFMT, and TAS score (Model 4,Table 5-2). The fully adjusted model (Model 4) showed that there was evidence for increased age being associated with poorer performance on the Bristol ERT, and that better performance on the GFMT was associated with better performance on the Bristol ERT (Table 5-2). There was no evidence that any of the other factors were associated with performance on the Bristol ERT. Collinearity of the variables included in the fully adjusted model (Model 4) was investigated and the variance inflation factor did not exceed two for any of the variables, suggesting that no correction was required.

5.3.4.1 Sensitivity analysis

Based on the information about TBI provided by participants the majority are classified as mild TBI given that they reported having loss of consciousness for 30 minutes or less. To check that the sample can be considered as mild TBI, the primary analysis was repeated with the eight participants who had or may have had loss of consciousness longer than 30 minutes excluded (Appendix C, section 8.15). The pattern

of results observed was no different, if anything, the trend towards decreased emotion recognition after TBI for the unadjusted model (Model 1, Table 8-12) became slightly stronger. Consequently, the whole TBI sample was used in the subsequent analysis and was considered representative of mild TBI.

Table 5-2

Regression models for overall emotion recognition on the Bristol ERT

Predictor	*b*	95% CI		*t*	*p*	Model Fit
		LL	UL			
Model 1						n = 110
(Intercept)	0.68	0.660	0.704			R^2 = .034
TBI	-0.03	-0.062	0.001	-1.94	0.05	95% CI [.00,.12]
Model 2						
(Intercept)	0.677	0.652	0.703			n = 110
TBI	-0.034	-0.066	-0.002	-2.08	0.04	R^2 = .040
DASS Anxiety score	0.001	-0.001	0.003	0.80	0.43	95% CI [.00,.12]
Model 3						
(Intercept)	0.706	0.613	0.800			
TBI	-0.018	-0.054	0.019	-0.97	0.34	
DASS Anxiety score	-0.000	-0.003	0.002	-0.28	0.78	
Sex (Female)	0.014	-0.021	0.049	0.82	0.41	n = 110
Age	-0.002	-0.003	-0.001	-3.55	<.001	R^2 = .163
Age left education	0.002	-0.001	0.005	1.15	0.25	95% CI [.03,.26]
Model 4						
(Intercept)	0.615	0.461	0.771			
TBI	-0.016	-0.051	0.019	-0.90	0.37	
DASS Anxiety score	0.001	-0.002	0.003	0.69	0.49	
Sex (Female)	0.003	-0.031	0.038	0.21	0.83	
Age	-0.002	-0.003	-0.001	-3.34	0.001	
Age left education	0.001	-0.001	0.005	0.93	0.35	n = 110
GFMT (total hits)	0.004	0.001	0.007	2.55	0.01	R^2 = .232
TAS (total score)	-0.001	-0.002	0.001	-1.17	0.25	95% CI [.06,.32]

Note: TBI is Traumatic Brain Injury. DASS is Depression, Anxiety, and Stress Scale. GFMT is Glasgow Face Matching Test. TAS is Toronto Alexithymia Scale. b is the unstandardised regression coefficient from a multiple linear regression. LL is the lower limit and UL the upper limit of the 95% Confidence Interval. t is the t-test coefficient.

Model 1 - Association between TBI and overall emotion recognition on the Bristol ERT
Model 2 - Model 1 adjusted for anxiety using the DASS anxiety subscale score
Model 3 - Model 2 additionally adjusted for sex, age, and age left education
Model 4 – Model 3 additionally adjusted for total hits on the GFMT and TAS score

5.3.5 Secondary analyses

5.3.5.1 Emotion specific outcomes

The primary regression analysis was rerun using the unbiased hit rate for each emotion and number of responses for each emotion as the outcomes to investigate emotion recognition accuracy and response bias for each individual emotion. There was no evidence for any emotion specific associations between emotion recognition accuracy and TBI or response bias and TBI in this sample (Table 5-3). There was also no evidence that anxiety was associated with emotion specific outcomes based on the anxiety coefficients from model 2 of the analysis (Appendix D, section 8.16).

5.3.5.2 Emotion Bias Tasks

Same as for the Bristol ERT a multiple linear regression was used to investigate differences between the TBI and control group on the two Emotion Bias Tasks (EBTs). The first model only included presence of TBI, and the second model was adjusted for anxiety, sex, age, age participants left education, performance on the GFMT, and TAS score. There was no evidence for a difference in balance point between the TBI and control group on either EBT; Angry (*b* = 0, 95% CI [-0.463, 0.463], t(108) = 0, p = 1.00) and Sad (*b* = 0.085, 95% CI [-0.401, 0.571], t(108) = 0.35, p = 0.73). There was no change when the model was adjusted for other factors; Angry (*b* = -0.08, 95% CI [-0.649, 0.498], t(102) = -0.26, p = 0.8), Sad (*b* = -0.086, 95% CI [-0.684, 0.513], t(102) = -0.28, p = 0.78). Further, there was no evidence that any of the factors included in the model were associated with performance on either EBT.

5.3.5.3 Other factors

There was weak evidence that the total score on the DASS was associated with overall emotion recognition, even after sex, age, age participants left education, performance on the GFMT, and TAS score were adjusted for. Participants with a higher DASS score were better at the Bristol ERT (*b* = 0.001, 95% CI [0.00003, 0.0014], t(102) = 2.09, p = 0.04). This corresponded with higher scores on both the depression and the stress subscales of the DASS being associated with an increase in emotion recognition performance; depression (*b* = 0.002, 95% CI [0.0003, 0.003], t(102) = 2.46, p = 0.02) and stress (*b* = 0.001, 95% CI [0.0001, 0.004], t(102) = 2.11, p = 0.04). There was no evidence that performance on the BPAQ was associated with a change in emotion recognition on the Bristol ERT (*b* = 0.0003, 95% CI [-0.001, 0.001], t(102) = 0.80, p = 0. 43).

Table 5-3

Results from regression analysis investigating association between TBI and emotion recognition performance for each emotion on the Bristol ERT

Emotion	Model 1 - Unadjusted					Model 2 - Adjusted for Anxiety					Model 3 - Fully Adjusted				
	b	95% CI		*t*	*p*	*b*	95% CI		*t*	*p*	*b*	95% CI		*t*	*p*
		LL	UL				LL	UL				LL	UL		
UNBIASED HIT RATE (HU)															
Angry	-0.004	-0.011	0.004	-0.94	0.35	-0.004	-0.012	0.004	-1.05	0.30	0.002	-0.007	0.011	0.39	0.69
Disgust	-0.006	-0.014	0.001	-1.74	0.09	-0.005	-0.013	0.002	-1.36	0.18	-0.005	-0.014	0.004	-1.17	0.24
Fear	-0.009	-0.022	0.004	-1.34	0.18	-0.012	-0.026	0.002	-1.75	0.08	-0.004	-0.018	0.011	-0.52	0.61
Happy	0.002	-0.005	0.008	0.44	0.66	0.001	-0.006	0.008	0.27	0.78	0.002	-0.006	0.011	0.55	0.59
Sad	-0.002	-0.008	0.005	-0.52	0.61	-0.003	-0.010	0.004	-0.88	0.38	-0.005	-0.013	0.003	-1.24	0.22
Surprise	-0.005	-0.011	0.001	-1.64	0.10	-0.005	-0.011	0.001	-1.59	0.12	-0.004	-0.011	0.003	-1.11	0.27
BIAS SCORE (number of responses given for each emotion)															
Angry	0.64	-0.9	2.17	0.82	0.41	0.43	-1.16	2.01	0.53	0.59	0.77	-1.08	2.61	0.82	0.41
Disgust	0.95	-1.20	3.09	0.88	0.38	0.50	-1.70	2.70	0.45	0.65	1.23	-1.32	3.77	0.96	0.34
Fear	-0.64	-2.85	1.57	-0.57	0.57	-0.52	-2.82	1.77	-0.45	0.65	-1.50	-4.24	1.23	-1.09	0.28
Happy	-1.31	-3.60	0.98	-1.13	0.26	-1.01	-3.38	1.35	-0.85	0.40	-1.34	-4.16	1.49	-0.94	0.35
Sad	-0.11	-1.93	1.71	-0.12	0.91	-0.02	-1.91	1.87	-0.02	0.98	0.78	-1.47	3.04	0.69	0.49
Surprise	0.47	-1.61	2.56	0.45	0.65	0.62	-1.54	2.79	0.57	0.57	0.06	-2.49	2.62	0.05	0.96

Note: b is the TBI unstandardised regression coefficient from a multiple linear regression. LL is the lower limit and UL the upper limit of the 95% Confidence Interval. t is the t-test coefficient. Model 1 - Association between TBI and emotion specific outcomes on the Bristol ERT. Model 2 - Model 1 adjusted for anxiety on the DASS. Model 3 - Model 2 additionally adjusted for sex, age, age left education, GFMT, and TAS.

5.3.6 Exploratory analyses

Given the trend towards decreased emotion recognition after TBI and the current literature around differences in outcome after mild TBI if people experienced Post Traumatic Amnesia (PTA), an exploratory analysis was conducted comparing TBI participants with PTA and those without PTA to the control group separately. There was evidence that the participants who reported having PTA show decreased performance on the Bristol ERT compared to neurologically healthy controls, whilst there was no evidence for a difference between the TBI group without PTA and the neurologically healthy controls (Model 1, Table 5-4). However, same as in the primary analysis there was no evidence for a difference between the TBI groups and the neurologically healthy controls once the model was fully adjusted (Model 4, Table 5-4).

5.3.7 Reliability

Internal reliability of accuracy on the Bristol ERT was assessed by calculating split-half reliability (Parsons et al., 2019) and Cronbach's alpha. For the split-half reliability all the trials were treated as the same condition and 5000 random splits were used to establish a Spearman-Brown corrected reliability estimate. The confidence interval for Cronbach's alpha was also calculated using 5000 bootstrap samples. The standard deviations and spearman brown corrected internal reliability estimates for used to calculate the Standard Error of Measurement (SEM) and the Smallest Real Difference (SRD). See Chapter 2 section 2.1.3 for details. In the full sample of 110 participants the Spearman-Brown corrected estimate for all the conditions was r_{sb} = 0.68 (95% CI: 0.59 to 0.76) and Cronbach's alpha for the 96 trials included in the Bristol ERT was α = 0.8 (95% CI: 0.75 to 0.83). To check whether low intensity trials on the Bristol ERT impacted reliability they were removed. The Spearman-Brown corrected reliability estimate was higher at r_{sb} = 0.73 (95% CI: 0.64 to 0.8) and Cronbach's alpha for the 60 items version was approximately the same at α = 0.79 (95% CI: 0.74 to 0.83). The SEM in this sample was 4.55 ($8.05\sqrt{1-0.68}$), which makes the SRD 12.6 hits (2.77×4.55). The split half reliability estimates were also calculated for the 55 participants with TBI and 55 neurologically healthy control participants. The Spearman-Brown corrected reliability estimate was about the same in the TBI group compared to the control group (r_{sb} = 0.68, 95% CI: 0.55 to 0.79, and r_{sb} = 0.69, 95% CI: 0.56 to 0.8, respectively). The SEM for the TBI group was 4.52 and 4.4 for the control group meaning that the SRD for the groups were 12.5 and 12.2 hits respectively.

Table 5-4

Regression models for overall emotion recognition on the Bristol ERT with the TBI group split into participants with PTA and those without PTA.

Predictor	*b*	95% CI		*t*	*p*	Model Fit
		LL	UL			
Model 1						
(Intercept)	0.682	0.660	0.704			n = 110
TBI with PTA	-0.048	-0.087	-0.010	-2.47	0.01	R^2 = .054
TBI without PTA	-0.015	-0.052	0.023	-0.78	0.43	95% CI [.00,.14]
Model 2						
(Intercept)	0.676	0.651	0.702			
TBI with PTA	-0.053	-0.093	-0.013	-2.63	0.01	n = 110
TBI without PTA	-0.018	-0.056	0.020	-0.93	0.35	R^2 = .062
DASS Anxiety score	0.001	-0.001	0.003	0.94	0.35	95% CI [.00,.15]
Model 3						
(Intercept)	0.700	0.607	0.794			
TBI with PTA	-0.041	-0.086	0.003	-1.83	0.07	
TBI without PTA	-0.003	-0.042	0.036	-0.16	0.88	
DASS Anxiety score	-0.000	-0.002	0.002	-0.08	0.94	
Sex (Female)	0.009	-0.026	0.045	0.54	0.59	n = 110
Age	-0.002	-0.003	-0.001	-3.56	0.001	R^2 = .188
Age left education	0.002	-0.001	0.006	1.40	0.16	95% CI [.03,.28]
Model 4						
(Intercept)	0.609	0.455	0.763			
TBI with PTA	-0.035	-0.079	0.009	-1.58	0.12	
TBI without PTA	-0.005	-0.044	0.033	-0.25	0.80	
DASS Anxiety score	0.001	-0.001	0.003	0.74	0.46	
Sex (Female)	0.000	-0.034	0.035	0.02	0.98	
Age	-0.002	-0.003	-0.001	-3.35	0.001	
Age left education	0.002	-0.001	0.005	1.15	0.25	n = 110
GFMT (total hits)	0.004	0.001	0.007	2.47	0.02	R^2 = .247
TAS (total score)	-0.001	-0.002	0.001	-0.97	0.34	95% CI [.06,.33]

Note: TBI is Traumatic Brain Injury. PTA is Post Traumatic Amnesia. DASS is Depression, Anxiety, and Stress Scale. GFMT is Glasgow Face Matching Test. TAS is Toronto Alexithymia Scale. b is the unstandardised regression coefficient from a multiple linear regression. LL is the lower limit and UL the upper limit of the 95% Confidence Interval. t is the t-test coefficient. The reference group for the TBI with PTA and TBI without PTA regression coefficients are the neurologically healthy controls.

Model 1 - Association between TBI and overall emotion recognition on the Bristol ERT
Model 2 - Model 1 adjusted for anxiety using the DASS anxiety subscale score
Model 3 - Model 2 additionally adjusted for sex, age, and age left education
Model 4 – Model 3 additionally adjusted for total hits on the GFMT and TAS score

5.4 Discussion

Based on the information about TBI available from the brain injury screening the participants recruited for the study can be classified as having mild TBI, so results will be discussed in context of the mild TBI literature. The aim of this study was to establish whether there was evidence for an association between emotion recognition on the Bristol ERT and mild TBI in a community-based sample. The results indicated that there was weak evidence for decreased emotion recognition being associated with mild TBI, which is in line with the primary hypothesis for this study. However, contrary to our predictions there was no evidence for this association once age, sex, face perception and emotion processing were controlled for. There was no evidence of emotion specific differences in emotion recognition between the mild TBI and control groups. Although this does not match our hypothesis predicting deficits in emotion recognition across all six emotions, it does show that there were no emotion specific differences in this sample. There was also no evidence of emotion bias on the Bristol ERT or EBT tasks associated with mild TBI. Despite evidence for increased anxiety in the mild TBI group compared to controls there was no evidence to support our hypotheses that anxiety would be associated with decreased emotion recognition accuracy and emotion specific bias.

5.4.1 Mild TBI

Having a mild TBI was associated with worse outcomes on all the questionnaires included in this study. Despite all participants being at least 10 months since the TBI they reported higher levels of anxiety and stress on the DASS compared to the control group. There was also a trend towards increased feelings of depression on the DASS. Participants with mild TBI had higher alexithymia scores on the TAS compared to the healthy controls, with the difference being particularly pronounced on the Externally Orientated Thinking subscale. Finally, there was also weak evidence for increased feelings of aggression based on the BPAQ. Although there is no way of knowing whether these negative outcomes are the result of the mild TBI in this sample, the findings are congruent with the current literature on negative outcomes after mild TBI (Calvillo & Irimia, 2020). The distribution of male versus female participants across the mild TBI versus control group showed that males were more likely to have experienced a TBI in this sample. This is congruent with the increased incidence of having a TBI if you are male reported in the UK (Headway, 2018). Males have tendency to score higher than

females on both the TAS and BPAQ (Buss & Perry, 1992; Parker et al., 2003). It is possible that scores on those questionnaires were higher for the mild TBI group compared to controls because there were more males in that group.

5.4.1.1 Emotion recognition accuracy

The results from this study indicated that there was no association between emotion recognition and mild TBI, which broadly matches the findings from previous studies investigating emotion recognition and mild TBI (McLellan & McKinlay, 2013; Theadom et al., 2019).The only difference in results was that Theadom and colleagues (2019) reported that four years post injury participants with mild TBI were less accurate at identifying happy stimuli, which was not observed in this study. There are many possible explanations for this difference. The Bristol ERT and EET from TASIT use very different stimuli so are likely to have different task demands, which has been shown to impact findings (Hayes et al., 2020). It is also that the lack of association between mild TBI and emotion recognition is the result of selection bias. Recruitment through Prolific suggests that participants were comfortable using a computer and are looking for research opportunities. Furthermore, participants on Prolific are generally well educated with median level of education being a bachelor's degree (Peer et al., 2017), which is consistent with the average age that participants in the sample reported leaving education being around 21. Higher levels of education have been associated with better outcomes after mild TBI (Cnossen et al., 2017). This could mean that participants in this study were not representative of people experiencing difficulties after mild TBI.

Although there was no evidence of decreased emotion recognition after mild TBI in this sample it is worth considering whether emotion recognition accuracy decreases with injury severity. It is possible that the inclusion criteria of head injury with loss of consciousness was too mild an injury to result in emotion recognition deficits. Arguably, presence of Post Traumatic Amnesia is a better indicator of mild TBI (Gasquoine, 2020), at minimum it suggests a higher level of severity compared to mild TBI with only loss of consciousness. The exploratory analysis, taking into consideration presence of Post Traumatic Amnesia, in this sample showed that only participants with Post Traumatic Amnesia had lower emotion recognition compared to the neurologically healthy controls. It is important to note that there was no evidence of an association between mild TBI with Post Traumatic Amnesia and emotion recognition once age and face perception had been controlled for. The results add to the growing evidence that

mild TBI is not associated with deficits in emotion recognition that are observed after moderate to severe TBI.

5.4.1.2 Emotion bias

Despite the evidence for increased levels of anxiety and aggression in the mild TBI group there was no evidence of mild TBI being associated with emotion recognition bias on the Bristol ERT and EBT tasks this sample. This was surprising given the evidence for negative attribution bias after TBI linked to aggression (D. Neumann et al., 2021) and association between state anxiety and emotion bias on the anger Bristol EBT reported by Attwood and colleagues (2017). The TBI group in the study by Neumann, Sander, and colleagues (2021) consisted mainly of participants with moderate to severe TBI. It is possible that, like with emotion recognition accuracy, negative attribution bias for emotions after TBI is only present for more severe TBI cases. It is also possible that aggression mediates the relationship between TBI and negative attribution bias for emotions, and that there was no meaningful difference in levels of aggression between the mild TBI and control group in this study. The lack of evidence for an association between anxiety and emotion bias is discussed in the next section. Investigating association between aggression (or indeed other factors) and emotion bias outcomes independent of TBI was not within the scope of this study.

5.4.2 Associations between Emotion Recognition and Covariates

5.4.2.1 Anxiety

The DASS was chosen as a measure of anxiety for this study because it is a useful continuous measure of clinical anxiety (Crawford & Henry, 2003; Lovibond & Lovibond, 1995). The aim was to capture clinically significant levels of anxiety without diagnosis of an anxiety disorder, which are thought to be much more prevalent than a diagnosis of Generalised Anxiety Disorder after TBI (Osborn et al., 2016). Contrary to our predications there was no evidence that the DASS anxiety score was associated with changes in emotion recognition accuracy or bias. The effect sizes found previously for deficits in emotion recognition associated with Generalised Anxiety Disorder have been small (Plana et al., 2014) and evidence for associations with sub clinical anxiety has been mixed (Attwood et al., 2017; Dyer et al., 2022; Suslow et al., 2019). Results from this study suggest that general symptoms of clinical anxiety are not associated with changes in emotion recognition accuracy. Given that Attwood and colleagues (2017) found that

induced state anxiety was associated with deficits in emotion recognition on the Bristol ERT, it is possible that emotion recognition is only impaired in people experiencing acute anxiety. Similarly, negative attribution bias for emotions could be associated with state anxiety and not general symptoms of anxiety. Attwood and colleagues (2017) found evidence for a shift towards perceiving anger on the angry/happy EBT after inducing anxiety. Interestingly, Dyer and colleagues (2022) were only able to replicate this effect in participants with high trait anxiety. It is possible that participants reporting symptoms of anxiety would show impairments and bias in emotion recognition when they experience acute stress, but this was not captured at the time of testing. Further studies are needed to investigate this, and these complex interactions highlight the need to clearly define aspects of anxiety in future research.

5.4.2.2 Age, Sex, and Education

There was a strong association between age and performance on the Bristol ERT indicating that emotion recognition decreased with age in this sample, which is congruent with previous findings (Abbruzzese et al., 2019; Byom et al., 2019; Ruffman et al., 2008). Murphy and colleagues (2019) have argued that the observed association could be mediated by general cognitive ability or presence of mood disorders, but the results in the current study did not support this. General cognitive ability was not directly assessed, but there was clear evidence for a negative association between emotion recognition and age after adjusting for face perception ability. Furthermore, depression and stress were associated with an increase, not a decrease, in performance on the Bristol ERT. It is highly unlikely that increased prevalence of mood disorders in older adults would explain the negative association between age and emotion recognition. Given the increased prevalence of TBI in older populations (Maas et al., 2017; Peeters et al., 2017) the potential impact of age on performance will need to be carefully considered. It would be helpful to establish age specific normative data for the Bristol ERT in both neurologically healthy and TBI populations. Arguably, females have small advantage over males in emotion recognition accuracy (Hoffmann et al., 2010; Thompson & Voyer, 2014), but this association was not observed in the current study. It is possible that the sample size was not big enough to detect this effect, especially considering that the sample included more female than male participants. There was also no evidence that years of education was associated with emotion recognition performance, which could be due to high levels of education observed in this sample.

5.4.2.3 Face perception and Alexithymia

The GFMT was included in this study to check whether differences in performance on the emotion recognition task between participants with mild TBI and controls can be explained through general face perception ability (J. D. Henry et al., 2015). It was not surprising to find that performance on the face discrimination task was positively correlated to performance on the Bristol ERT given that facial stimuli are used in the Bristol ERT. The TAS mean and standard deviation for the control group in this sample was reflective of the normative data scores established in a Canadian population (Mean = 45.75, SD = 11.3; Parker et al., 2003). There was evidence that participants with mild TBI had higher scores on the TAS, but there was no evidence that performance on the TAS was associated with emotion recognition in the fully adjusted regression model. Notably, although the mean score for the mild TBI group was higher it was within the range described as low alexithymia (Bagby, Parker, et al., 1994; Bagby, Taylor, et al., 1994), which could explain why no association was observed in this sample. As discussed in Chapter 3, it is also possible that there simply is no association between accuracy on emotion labelling tasks and alexithymia (Ihme, Sacher, Lichev, Rosenberg, Kugel, Rufer, Grabe, Pampel, Lepsien, Kersting, Villringer, Lane, et al., 2014; Rosenberg et al., 2019). Studies reporting an association between the TAS and emotion recognition have found that the highest correlation was with the Externally Oriented Feelings subscale (Martinez-Sanchez et al., 2017; Prkachin et al., 2009). In the current study, participants with mild TBI scored higher that the controls participants on that subscale, so it would be worth investigating that specific interaction further (Neumann, Zupan, et al., 2014).

5.4.2.4 Depression, Stress, and Aggression

Contrary to expectations based on the literature (Dalili et al., 2015; Krause et al., 2021), there was evidence that higher levels of depression and stress measured using the DASS were associated with better performance on the Bristol ERT. A possible explanation is that the DASS is capturing general negative affect in the same way that the trait subscale of the State-Trait Anxiety Inventory (Spielberger, 1983) is thought to measure negative affect (Knowles & Olatunji, 2020). Attwood and colleagues (2017) found that higher trait scores on the State-Trait Anxiety Inventory were associated with increased performance on the Bristol ERT, which would correspond with the findings in this study. However, these findings should be interpreted with caution because the analyses were exploratory and evidence for the effect was not very strong and could be

a false positive. There was no evidence that aggression was associated with emotion recognition in this sample. Notably, the mean scores on the BPAQ for this study were lower than the mean scores for the normative sample (Buss & Perry, 1992), suggesting that low levels of aggression may account for the lack of association.

5.4.3 Reliability of the Bristol ERT

Internal consistency estimates were calculated for the Bristol ERT to assess the extent to which items consistently measured the same construct (Sherman et al., 2011). Using the thresholds for reliability estimates outlined by Strauss and colleagues (2006) the split half reliability for the Bristol ERT in this sample can be described as marginal for both the overall sample and the TBI and control groups separately. The fact that the reliability estimates were comparable across the mild TBI and control groups, suggests that having a mild TBI does not impact reliability of measurement on the Bristol ERT. The split half reliability estimate for the Bristol ERT increased when low intensity trials were removed, indicating that variance was increased due to the highly ambiguous low intensity trials. It is promising that the reliability estimate for the reduced Bristol ERT is within in the adequate range and meets the reliability threshold for tasks considered suitable for research (Ponterotto & Ruckdeschel, 2007; Strauss et al., 2006).

Cronbach's alpha is the most commonly reported measure of internal consistency reliability, which is why it was included alongside the split half reliability estimate recommended by Parson's and colleagues (2019). Unlike the split half estimate Cronbach's alpha can be described as high for the full Bristol ERT in this sample and was only slightly lower for the reduced Bristol ERT. This suggests that the split half method is more sensitive to inclusion of low intensity trials, potentially because it is based on 5000 random splits as opposed to an estimate of all possible splits like Cronbach's alpha. The level of reliability indicated by Cronbach's alpha suggests that measurements on the Bristol ERT could be sufficiently reliable for both research and individual assessment in a clinical setting (Sattler, 2001; Strauss et al., 2006). The Smallest Real Difference estimate shows that if the Bristol ERT were used to assess individual performance, a change in score of 13 hits would be needed to be confident of a true difference in performance. Wider implications in the context of reliability estimates associated with the Bristol ERT in the other studies presented in this thesis will be addressed in the general discussion.

5.4.4 Strengths and Limitations

The current study is one of only few studies that has investigated the association between mild TBI and emotion recognition. A benefit of the study design was that well validated measures were included to assess several potential confounders alongside the Bristol ERT. A possible limitation was that measures of general cognitive ability, for example working memory and processing speed were not included due to constraints around study duration. Instead, a measure of face identification was included as recommended by Henry and colleagues (2015) when assessing emotion recognition using facial stimuli. Another potential limitation was that the sample size calculation for this study was based on the effect size for emotion recognition deficits after moderate to severe TBI. Consequently, it is possible that the sample was simply not big enough to detect smaller effects. Furthermore, the reliability coefficients for the Bristol ERT suggested that variance in scores could be quite high, which decreases confidence in the inferences made in this study. A concern is that online data collection could have impacted quality of the data collected, which in turn would have impact reliability estimates. However, research has shown that data collected using Prolific is generally of high quality (Peer et al., 2017) and results from online data collection are often comparable to results using more conventional forms of data collection (Walter et al., 2019).

5.4.5 Implications and future directions

Ultimately this study did not find any evidence to support an association between emotion recognition and mild TBI. It is possible that using head injury with loss of consciousness is too broad a category for mild TBI. Future studies should consider how mild TBI is defined and whether different subgroups, such as those with Post Traumatic Amnesia, are associated with different outcomes on emotion recognition tasks. The findings also highlight the impact of age on emotion recognition, which will need to be taken into consideration if the task is used for individual assessments in performance. It would be useful to establish age specific normative data for the Bristol ERT to use as a benchmark for both research purposes and clinical practice.

Chapter 6 Associations between mild Traumatic Brain Injury, emotion recognition, and anxiety in the Avon Longitudinal Study of Parents and Children

6.1 Introduction

In Chapter 5 the association between mild TBI and emotion recognition was assessed using a cross sectional study design. The study in this chapter aimed to address the same research questions in a prospective cohort study. The primary aim was to investigate the associations between mild TBI and emotion recognition performance on the Bristol ERT in the Avon Longitudinal Study of Parents and Children (ALSPAC). The secondary aim was to explore whether anxiety is associated which changes in emotion recognition in this cohort to help evaluate whether it could be moderating changes in emotion recognition in a mild TBI population. Triangulation across studies will help address biases associated with a single research methodology and further develop our understanding of associations between emotion recognition, mild TBI and anxiety.

6.1.1 Observational research: Prospective cohort studies

Data for the current study was extracted from a prospective cohort study, which is a longitudinal study where participants are recruited and then followed up over a period of time (Grimes & Schulz, 2002). Prospective cohort studies can reduce selection bias because participants are not recruited retrospectively from a population that has already experienced a TBI. However, cohort studies are susceptible to other sources of selection bias. Participants with certain exposures are more likely to drop out and thus bias observations made based on the cohort data (Munafo et al., 2018). For example, participants with complications after TBI might be more likely to drop out. This could mean the true extent of difficulties mild TBI are not captured because participants presenting with difficulties are no longer part of the study. In order to minimise impact of bias it is important to consider and control for possible confounding factors during analysis of cohort studies (Bell, 2020). A benefit of cohort studies is that there is data for many variables, so it is possible to control for a lot of confounders and consider multiple outcomes for a given exposure such as mild TBI. Another advantage of a cohort study is that data is available for large number of participants. This means that that is possible to detect small effects that could otherwise be masked, which is important given that changes in emotion recognition after mild TBI could be subtle.

6.1.1.1 Avon Longitudinal Study of Parents and Children (ALSPAC)

As outlined in Chapter 2 section 2.1.6, ALSPAC is a birth cohort of initially 14,541 pregnancies. Pregnant women resident in Avon, UK with expected dates of delivery 1st April 1991 to 31st December 1992 were invited to take part in the study. There was a total of 14,676 foetuses, resulting in 14,062 live births and 13,988 children who were alive at 1 year of age. When the oldest children were approximately 7 years of age, an additional 913 children not included in the initial phase were enrolled. Details about the phases of enrolment and further information about the cohort are described in the cohort profile papers and update (Boyd et al., 2013; Fraser et al., 2013; Northstone et al., 2019). The total sample size for analyses using any data collected after the age of seven is 15,454 pregnancies, resulting in 15,589 foetuses. Of these 14,901 were alive at 1 year of age. Data about the children's, health, development, and well-being was initially collected using parent completed questionnaires, as well as clinics attended by the children from around age 7, and later on child completed questionnaires. Further information about the cohort profile is available at https://www.bristol.ac.uk/alspac/.

6.1.2 Current study: Rationale

As has been discussed previously (section 5.1.1), there are very few studies that have investigated the associations between emotion recognition and mild TBI (Calvillo & Irimia, 2020). Theadom and colleagues (2019) did not find evidence for a difference in overall emotion recognition accuracy between people with mild TBI and healthy controls and week evidence for lower accuracy for happy stimuli. Participants for their study were recruited from the Brain Injury Incidence and Outcomes New Zealand (NZ) in the Community, BIONIC study. Any participants who were 12 years or older at the time of injury were included. Usually, studies of adult TBI include only participants over the age of 16 (Babbage et al., 2011; Murphy et al., 2021) because emotion recognition is considered fully developed by age 16 (Lawrence et al., 2015). Having a TBI before emotion recognition is fully developed could impact emotion recognition differently to a TBI that is received when emotion recognition is fully developed (Giza & Prins, 2006). However, a study looking at impact of childhood TBI on adult social cognition found no differences in emotion recognition accuracy between participants with mild TBI and orthopaedic controls (McLellan & McKinlay, 2013). Notably, the emotion recognition task used showed faces only at 100% intensity and sample sizes were small, so they may not have been able to pick up small changes in emotion recognition after mild TBI.

In the current study, the primary aim was to explore the impact of mild TBI occurring after emotion recognition is considered fully developed. Adult mild TBI was defined as having had a head injury with loss of consciousness at age 17 or later. As prior studies have also included childhood TBI the secondary TBI variable was created including all participants in ALSPAC who have had a mild TBI. Maximising the mild TBI sample in this way could help detect small changes in emotion recognition associated with having a mild TBI. Emotion recognition for all participants was measured using the Bristol ERT when participants were around 24 years of age.

6.1.2.1 Anxiety as a potential moderator

There is evidence that the likelihood of having an anxiety disorder is increased after mild TBI (Barker-Collo et al., 2018; Delmonico et al., 2021), which could in turn be impacting performance on emotion recognition tasks. Anxiety disorders, including Generalised Anxiety Disorder (GAD) have been associated with deficits in emotion recognition accuracy (Plana et al., 2014). Palm and colleagues (2011) investigated changes in emotion recognition and brain functioning in a group of 15 females with GAD. They reported decreased recognition of sad stimuli and decreased activation of brain areas associated with emotion recognition for all emotion stimuli. The results should be interpreted with caution given the small sample size, but it does suggest that increased prevalence of GAD after mild TBI could impact performance on emotion recognition tasks. In the current study, the associations between emotion recognition, presence of GAD and reports of anxiety in the week before completing the Bristol ERT were investigated. Of interest was the potential moderating effect of anxiety on the association between emotion recognition and mild TBI in the ALSPAC cohort. We used two different definitions of anxiety to contrast the impacts of having an anxiety disorder and sub clinical reports of anxiety.

6.1.3 Current study: Aims and Hypotheses

The associations between emotion recognition, mild TBI and anxiety in ALSPAC were investigated in a series of steps. Given the lack of evidence for changes in emotion recognition after mild TBI and conflicting evidence regarding links between anxiety and emotion recognition the analysis was considered explorative rather than confirmative. All the hypotheses were tentative, and no emotion specific hypotheses were made. The first step was to investigate the association between emotion recognition and mild TBI in this cohort. Our hypothesis was that participants with mild TBI might perform slightly

worse on the Bristol ERT compared to control participants. The second step was to investigate whether mild TBI was associated with GAD in this sample. Our hypothesis was that mild TBI would be associated with an increased likelihood of GAD even after controlling for presence of GAD preinjury. The third step was to investigate the association between emotion recognition and concurrent anxiety. Both diagnosis of GAD and presence of anxiety in the week before completing the Bristol ERT were considered. The hypothesis was that both GAD and recent anxiety would be associated with decreased emotion recognition on the Bristol ERT. The fourth and final step was to consider whether anxiety could be mediating changes in emotion recognition after mild TBI in this sample. Our hypothesis was that if there are changes in emotion recognition after mild TBI, they could be mediated by anxiety at the time when the emotion recognition task was completed. Participants with mild TBI were compared to participants who had experienced an orthopaedic injury and participants with no reported injury. Age, sex, social class, and general cognitive ability were considered as potential confounders and controlled for in the analyses.

6.2 Method

6.2.1 Design

Associations between mild Traumatic Brain Injury reported between 17 and 25 years of age, anxiety at age 24 and emotion recognition at age 24 are investigated using secondary data from a prospective cohort study. The two main outcomes for this study are overall emotion recognition on the Bristol ERT and presence of Generalised Anxiety Disorder at age 24. Secondary outcomes are emotion recognition accuracy for each of the six emotions and the number of responses for each emotion. Ethical approval for the study was obtained from the ALSPAC Ethics and Law Committee and the Local Research Ethics Committees.

6.2.2 Participants

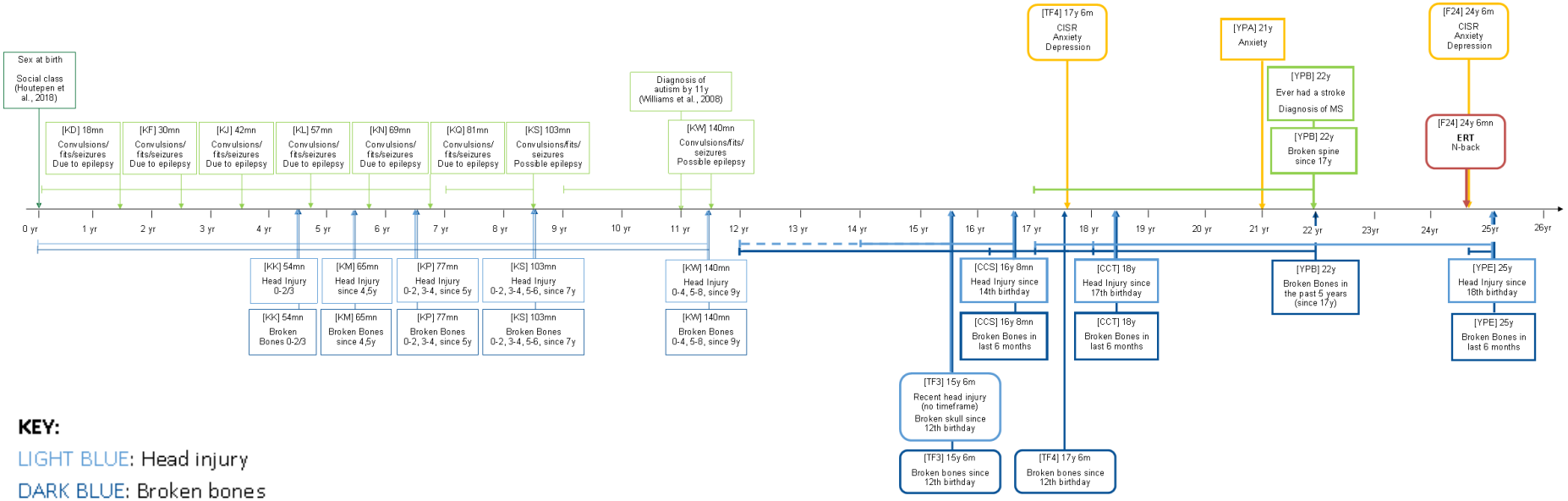
The participants were drawn from the Avon Longitudinal Study of Parents and Children (ALSPAC). At around age 24 the children in this cohort (now young adults) were invited to attend a clinic at which they completed the Bristol ERT as a measure of emotion recognition and the Clinical Interview Schedule Revised (Lewis et al., 1992) was used to assess anxiety. The clinic was held between June 2015 and October 2017, with 4026 of the young people attending the clinic during that time. Information about adult mild Traumatic Brain Injury was obtained from questionnaires completed around age 18 and age 25.

6.2.3 Measures

The measures included in this study were a combination of information collected at clinic and questionnaires completed by the children/young adults or their parents. See Figure 6-1 for a timeline including all the timepoints at which data was collected for the variables included in this study. At clinic, the study data were collected and managed using REDCap electronic data capture tools hosted at the University of Bristol. REDCap (Research Electronic Data Capture) is a secure, web-based software platform designed to support data capture for research studies (Harris et al., 2009). The cognitive tasks were presented using E-prime software (PST Inc, Sharpsburg, PA, USA). Please note that the study website contains details of all the data that is available through a fully searchable data dictionary and variable search tool (http://www.bris.ac.uk/alspac/researchers/data-access/data-dictionary/).

Figure 6-1

Timeline for ALSPAC variables



KEY:

LIGHT BLUE: Head injury

DARK BLUE: Broken bones

RED: Cognitive tests

YELLOW: Mood questions

DARK GREEN: Sociodemographic information

LIGHT GREEN: Exclusions

Shapes:

Rectangle = questionnaire based

Rounded rectangle = clinic based

6.2.3.1 Emotion recognition

The Bristol ERT was used as a measure of emotion recognition at the clinic that occurred around age 24. The task consisted of 96 trials including six emotions at eight levels of intensity for a set of Caucasian male and a set of Caucasian female faces. The order of presentation of the stimuli was randomised for each participant and stimuli were presented for 200ms. The primary outcome was overall emotion recognition using the proportion of total hits for each participant, so the number of correct responses divided by the number of trials completed. The secondary outcomes were emotion specific accuracy and response scores measured using unbiased hit rate for each of the six emotions and the total number of responses made for each emotion respectively. For a detailed description of the task and outcomes please refer to the introduction chapter section 1.3.2.

6.2.3.2 Mild Traumatic Brain Injury

Information about mild Traumatic Brain Injury (mild TBI) was available from questionnaires asking whether participants have experienced head injury with loss of consciousness. In a questionnaire at around age 25 participants were asked whether they have had a head injury with loss of consciousness since their 18th birthday, and at around age 18 they were asked whether they have had a head injury with loss of consciousness since their 17th birthday. Responses on these two questions were combined to create a measure of adult mild TBI defined as head injury with loss of consciousness between ages 17 and 25. To ensure that childhood mild TBI did not impact results, any participants who had a head injury with loss of consciousness between birth and age 16 were excluded. The mild TBI variable for injury between birth and age 16 is based on a variable created previously (Kennedy et al., 2017) and consists of questionnaire responses at 7 timepoints asking about head injury with loss of consciousness or a broken skull (Figure 6-1). In addition to the adult mild TBI variable a lifetime mild TBI variable was created by combining the birth to age 16 variable and the 17 to age 25 variables. The number of times a participant reported a mild TBI was not considered given the information available about TBI in this cohort.

Two control group variables were generated to use for comparison with the mild TBI group. The first is a negative control of participants who reported a broken bone within a timeframe comparable to the mild TBI. Using orthopaedic injury as a

comparison group should control for distress caused simply by having a traumatic injury (Curran et al., 2000; Kennedy et al., 2017). Participants who reported a broken skull and those with both a mild TBI and broken bones were included in the mild TBI group. Further, participants who reported breaking fingers, toes or similar minor breaks were not included in the broken bones group, and participants who reported a broken spine were excluded from the study as this was deemed too severe an injury to be comparable to a mild TBI. This means that the broken bones group consisted of participants who reported breaking an arm, leg, ankle, or other bone. Information about broken bones between 17 and 25 was available at two time points, which were combined to create a broken bones group for that timeframe. At around age 22 participants were asked about broken bones since their 17th birthday and at around age 25 they were asked about broken bones in the past 6 months. No information about broken bones was available for the participants between the ages 22 years and 24 years 6 months. The broken bone variable for birth to 16 years was based on a broken bone variable from birth to 11 years created previously (Kennedy et al., 2017), and responses about broken bones at 15 years 6 months and 16 years 8 months. A lifetime broken bones variable was created by combining information across all available timepoints.

The second control group were participants with no injury. Participants who responded NO to having a head injury and NO to having an accident that led to a broken bone at four timepoints after the age of 17 (see Figure 6-1) were included in the no injury group from 17 to 25 years. Further, any participants with head injury between birth and 16 years were removed from this group; having a broken bone prior to 17 was not deemed a reason for exclusion. A lifetime no injury group was a combination of a no injury group between birth and 16 created previously (Kennedy et al., 2017) and the no injury group between 17 and 25 group created for this study.

6.2.3.3 Anxiety

The computerised version of the Clinical Interview Schedule – Revised (CIS-R) was used to identify presence of anxiety disorders according to diagnostic criteria in the ICD-10 (Lewis et al., 1992; Patton et al., 1999). Participants completed the self-administered computerised version of the interview at clinic around the age of 17 years and 6 months, and then again at around age 24. Responses were coded into diagnosis variables by the ALSPAC team and made available alongside responses for each individual item. For this study participants who met diagnostic criteria for Generalised

Anxiety Disorder (GAD) were included in the anxiety group. Participants who did not meet any diagnostic criteria for a mood disorder on the CIS-R (i.e. were not identified has having any anxiety disorders or depression) were used as a control group. The primary outcome was a binary variable indicating presence of GAD at age 24 based on these groups. Similarly, a binary variable for presence of GAD at age 17 years and 6 months was generated to control for anxiety prior to injury. As part of the CIS-R, participants responded to the following question: "On how many of the PAST SEVEN DAYS have you felt GENERALLY anxious, nervous or tense?" Responses to this question were used to create a secondary binary outcome variable indicating whether participants had experienced general anxiety in the week prior to attending the clinic at age 24.

6.2.3.4 Confounding variables

Sex, age at measure completion, and socioeconomic status were all included as confounders in this study to control for potential selection bias. The sex variable was binary based on the sex participants were assigned at birth and an age variable was calculated using the age in months reported at the clinical around age 24 divided by 12. Maternal social class at the birth of their child was used as a proxy for young person socioeconomic status in this study. A maternal social class variable in ALSPAC has been derived from maternal self-reported highest occupation level related to the Registrar General's classification of occupations. The categorisation ranges from social class V (unskilled manual) to social class I (professional), but for this study a binary variable was created with social classes IV and V in one group and social classes I to III and armed forces in the other.

At the same time as completing the emotion recognition task at around age 24, participants completed a measure of working memory. To adjust for potential changes in cognitive ability after mild TBI, working memory performance was included as a confounding variable in this study. A visuoverbal 2-back version of the N-back task was presented using E-prime (Kirchner, 1958). Please refer to Chapter 2 on reliability of tasks used in ALSPAC for a detailed description of this task. The outcome measure used was a discriminability index (*d*'prime) as a measure of overall performance on the task. As described above, we also adjusted for anxiety at age 17 (prior to the mild TBI occurrence) to try and ensure that TBI preceded changes in anxiety.

6.2.3.5 Exclusions

A diagnosis of autism was used as an exclusion criterion as there is strong evidence for altered emotion recognition in people with autism (Uljarevic & Hamilton, 2013). A binary variable indicating a diagnosis of autism before the age of 11 was created by Williams et al. (2008) and was used to exclude individuals with autism. Reports of neurological disorders other than mild TBI were also used as an exclusion criterion because they could impact emotion recognition performance at a later stage. Stroke can cause brain damage and has been shown to cause deficits in emotion recognition (Yuvaraj et al., 2013) and Multiple Sclerosis has also been linked to deficits in emotion recognition (A. Henry et al., 2015; Prochnow et al., 2011). At around age 22 years, participants in ALSPAC completed a questionnaire asking whether they had ever had a stroke or been given a diagnosis of Multiple Sclerosis. Any participants who indicated that they had were excluded from this study. Epilepsy in certain areas of the brain has also been associated with decreased emotion recognition (Bonora et al., 2011). Having a head injury is associated with a higher risk of developing epilepsy, but it is difficult to differentiate between post-traumatic epilepsy and focal epilepsy without injury (Fordington & Manford, 2020). Consequently, any participants who reported having had a seizure likely due to epilepsy prior to age 12 were excluded. Information about seizures likely due to epilepsy was available through parent report at 7 timepoints between birth and 11 years and 6 months (Figure 6-1). An epilepsy variable was generated by combining responses across those times points. Reporting a broken spine was also used as an exclusion criterion as the severity of the injury was likely to have long term effects that do not match the broken bones category or mild TBI category. For all these variables, participants with missing data were not excluded from the analysis.

6.2.4 Analysis Plan

To investigate the associations between mild Traumatic Brain Injury (mild TBI), emotion recognition, and anxiety in the ALSPAC cohort a stepped approach was used to investigate the associations between mild TBI and emotion recognition, mild TBI and anxiety, as well as emotion recognition and anxiety independently. A mediation analysis was planned as a final step, but this was not deemed appropriate given the results of the first three analysis steps and data available after extraction. The relevant variables were extracted from the ALSPAC dataset using Stata version 16 (StataCorp LLC), which was also used to generate the main variables used in this study. The data was then loaded

into RStudio (2020) R version 4.0.4 for data cleaning and analysis. The main analysis packages used were the core *stats* package in R, *psych* package version 2.1.9, and *apaTables* package version 2.0.5 (Stanley, 2018).

6.2.4.1 Step 1 – Association between mild TBI and emotion recognition at age 24

The first step of the analysis was to investigate whether having a mild TBI between ages 17 and 25 is associated with a change in emotion recognition compared to participants who had not reported an injury and those who had reported breaking an arm/hand or leg/foot (negative control group). A linear multiple regression was used with emotion recognition performance on the Bristol ERT as the outcome and injury as a categorial exposure variable. The reference group was changed from no injury to broken bones to obtain comparisons between all three injury groups. The model was adjusted, first for demographic variables (age at the time the outcome measure was completed, sex at birth, and maternal social class), and then further adjusted for working memory performance. The same analysis was used to investigate whether having a mild TBI at any age (birth to 25 years) resulted in a change in emotion recognition compared to participants with no injury or those who reported broken bones. To investigate emotion specific changes in performance on the Bristol ERT the above analysis for injury between 17 and 25 years was repeated using first unbiased hit rate for each emotion and then response bias scores (number of responses for each emotion) as the outcome measures. Finally, a sensitivity analysis was conducted with potential outliers on the Bristol ERT removed. An outlier was defined as scores more than 1.5 times the interquartile range above the 3rd or below the 1st quartile on the primary outcome for the Bristol ERT.

6.2.4.2 Step 2 – Association between mild TBI and anxiety at age 24

The second step of the analysis investigated whether reporting a mild TBI between 17 and 25 changes the likelihood of having an anxiety disorder at age 24. A logistic regression was conducted with presence of GAD at age 24 as to outcome and injury between 17 as the exposure. The analysis was adjusted for demographic variables (age at time of the outcome measure, sex at birth, and maternal social class) as possible confounders, and then adjusted for presence of GAD at 17 years and 6 months because it preceded time of injury in most cases. Odds ratios were calculated to show the likelihood of having GAD in the mild TBI group compared to the no injury group. The

same analysis was conducted using the lifetime injury group, but the analysis was not adjusted for anxiety at age 17 and 6 months.

6.2.4.3 Step 3 – Association between anxiety and emotion recognition at age 24

The third step of the analysis investigated whether having anxiety at age 24 was associated with performance on the Bristol ERT at age 24. Multiple linear regression with emotion recognition performance on the Bristol ERT as the outcome and presence of GAD as a binary exposure was conducted. Same as in step 1 the model was adjusted for demographic variables (age at time of the outcome measure, sex at birth, and maternal social class) and working memory performance on the n-back task at age 24. The analysis was also run using the binary variable indicating whether participants had experienced anxiety in the past week as the exposure, instead of presence of GAD. To check for emotion specific associations, both the analysis using the GAD variable, and analysis using the anxiety in the past week variable were repeated using unbiased hit rate and response bias scores for each emotion as the outcome variables.

6.3 Results

6.3.1 Sample available for analysis

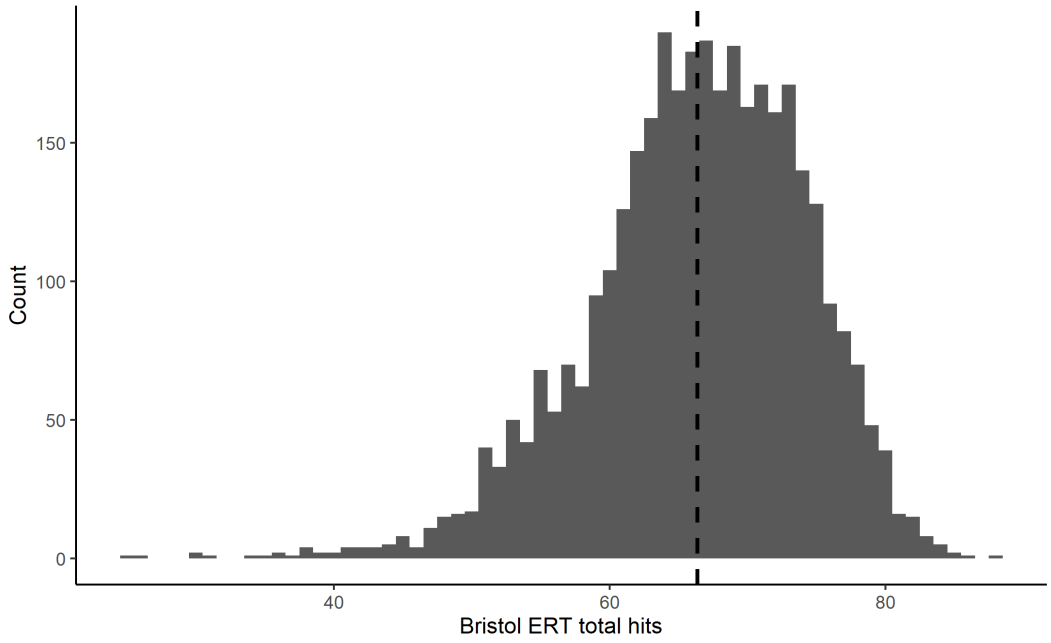
Of the 4026 participants who attended the clinic at age 24 only participants with the required outcome data were included in the analysis. This means in steps 1 and 3 of this analysis the 3551 participants who completed all 96 trials on the Bristol ERT were included. In step 2 the 3639 participants with diagnostic information about Generalised Anxiety Disorder (GAD) available were included. Sample sizes were further reduced due to missing exposure data and missing covariate data. The number of participants in each exposure group for every step of the analysis is included in Appendix A (section 8.17).

6.3.2 Properties of the Bristol ERT in ALSPAC

The mean performance on overall emotion recognition in the full sample of 3551 participants who completed the Bristol ERT is 66.36 (SD = 7.87), with a range of 28 to 88 total hits. Shapiro-Wilk test for normality indicated that the data was not normally distributed ($W(3550) = 0.98$, $p < .001$), however given the size of the dataset analysis was not adjusted based on the violation of normality (Schmidt & Finan, 2018).

Figure 6-2

Histogram for Bristol ERT total hits in ALSPAC



Notes: The histogram shows the number of times a particular score was obtained on the Bristol ERT. The dashed line represents the mean score on the Bristol ERT.

6.3.3 Step 1 – Association between mild TBI and emotion recognition at age 24

6.3.3.1 Descriptive statistics

Of the 3551 participants who completed the Bristol ERT at around age 24 only participants with injury data available were included in this analysis, n = 1146 for injury between ages 17 and 25, and n = 1903 for injury between birth and age 25. Descriptive statistics for participants with injury data between ages 17 and 25 are presented in Table 6-1 and for participants with injury data from birth to age 25 in Table 6-2.

Table 6-1

Descriptive statistics for primary outcome & confounders by injury between ages 17 & 25

Variable	Total		No Injury		mild TBI		Broken Bones		Injury group comparison	
	n = 1146		n = 748		n = 254		n = 144			
Categorical	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**χ^2**	***p***
Female	799	70%	558	75%	154	61%	87	60%	24.3	<.001
Social Class IV & V	61	6%	44	7%	10	5%	7	6%	1.0	0.61
Continuous	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	***F-value***	***p***
Age at clinic (years)	24.4	0.7	24.3	0.7	24.4	0.7	24.5	0.7	3.9	0.02
N-back (d-prime)	2.75	0.9	2.74	0.9	2.78	0.9	2.74	0.8	0.17	0.84
Bristol ERT (total hits)	66.9	7.7	67.2	7.8	66.8	7.2	65.7	8.3	2.3	0.10

Note: Sample size can vary due to missing covariate data. Social class variable is maternal social class. Continuous variables were all completed at clinic around 24 years. Comparisons are chi-squared tests for the categorial variables and analysis of variance for the continuous variables.

The proportion of participants in each group for the categorical variables, as well as the means and standard deviations for the continuous variables are comparable to the injury between the adult injury (ages 17 to 25) and the lifetime injury (birth to age 25) data sets. The results from the group comparisons indicate that there was a higher proportion of males in the mild TBI and broken bones group compared to the no injury group, suggesting that males are more likely to have an injury than females. There are very few participants in the lower social class category, but the proportion of these participants was the same across the injury groups. There was evidence for an effect of age across the groups. However, there was little variation in age within this sample (min = 22.75, max = 26.25) and the mean age was similar across groups. Large sample sizes are sensitive to small differences and this age effect is unlikely to be a source of bias.

Table 6-2

Descriptive statistics for primary outcome & confounders by injury between birth & 25

Variable	Total		No Injury		mild TBI		Broken Bones		Injury group comparison	
	n = 1903		n = 612		n = 530		n = 761			
Categorical	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**χ^2**	***p***
Female	1209	64%	465	76%	298	56%	446	59%	61.1	<.01
Social Class IV & V	110	7%	35	7%	30	7%	45	7%	0.1	0.95
Continuous	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	***F-value***	***p***
Age at clinic (years)	24.4	0.8	24.3	0.7	24.4	0.7	24.5	0.8	10.2	<.01
N-back (d-prime)	2.74	0.9	2.73	0.9	2.75	0.9	2.73	0.9	0.09	0.92
Bristol ERT (total hits)	66.6	7.7	67.1	7.9	66.3	7.4	66.4	7.7	1.96	0.14

Note: Sample size can vary due to missing covariate data. Social class variable is maternal social class. Continuous variables were all completed at clinic around 24 years. Comparisons are chi-squared tests for the categorial variables and analysis of variance for the continuous variables.

6.3.3.2 Overall emotion recognition accuracy

A multiple linear regression with proportion of total hits as an outcome and injury as the exposure was used to analyse the association between emotion recognition and injury in both the adult and lifetime injury groups (Table 6-3). See Appendix A (Section 8.17) for the number of participants in each injury group for each model.

There was no evidence that having a mild TBI between ages 17 and 25 impacts overall emotion recognition at age 24 in this sample (β = -0.004, 95% CI [-0.015, 0.008], *t*(1143) = -0.64, *p* = .52). There was weak evidence for decreased emotion recognition after breaking a bone in that same time period (β = -0.016, 95% CI [-0.030, -0.001], *t*(1143) = -2.13, *p* = .03), but there was no evidence of this association once age, sex, and maternal social class have been controlled for. Further, there was no evidence for a difference in overall emotion recognition performance between participants with mild TBI and participants with a history of broken bones (β = 0.012, 95% CI [-0.005, 0.028], *t*(1143) = 1.41, *p* = .16). Similarly, there was no evidence that lifetime injury of mild TBI or broken bones was associated with overall emotion recognition at age 24. See Table 6-3 for results of injury comparisons from all three models.

Table 6-3

Results from regression analysis investigating impact of injury on overall emotion recognition

Time of Injury		Unadjusted					Model 1					Model 2				
		n	Estimate	95% CI		p	n	Estimate	95% CI		p	n	Estimate	95% CI		p
				LL	UL				LL	UL				LL	UL	
Age 17 to 25	mild TBI vs None	1,146	-0.004	-0.015	0.008	0.52	940	-0.001	-0.013	0.012	0.89	918	0.000	-0.012	0.013	0.94
	BB vs None		-0.016	-0.030	-0.001	0.03		-0.011	-0.027	0.005	0.17		-0.008	-0.024	0.007	0.30
	mild TBI vs BB		0.012	-0.005	0.028	0.16		0.010	-0.008	0.028	0.27		0.009	-0.009	0.026	0.34
Birth to age 25	mild TBI vs None	1,903	-0.008	-0.018	0.001	0.08	1,569	-0.003	-0.013	0.007	0.61	1,538	-0.001	-0.011	0.009	0.84
	BB vs None		-0.007	-0.016	0.001	0.09		-0.004	-0.013	0.005	0.39		-0.001	-0.010	0.008	0.85
	mild TBI vs BB		-0.001	-0.010	0.008	0.81		0.001	-0.008	0.011	0.78		0.000	-0.009	0.009	0.98

Note: Sample size reduces for each model as participants who are missing covariate data used in that model are excluded. mild TBI refers to mild Traumatic Brain Injury. BB refers to Broken Bones. None refers to No Injury. Estimate is regression coefficient from a multiple linear regression. LL is the lower limit and UL the upper limit of the 95% Confidence Interval. In bold are the estimates with evidence for an association between injury and emotion recognition.

Models:
Unadjusted: Impact of receiving an injury on emotion recognition performance measured using the Bristol ERT at age 24
Model 1: As Unadjusted and additionally adjusted for age at 24 clinic, sex at birth, and maternal social class
Model 2: As Model 1 and additionally adjusted for working memory (2-back) at age 24

The results from the fully adjusted model testing the association between overall emotion recognition and adult injury indicate that the sex, social class, and working memory covariates were associated with emotion recognition performance. There was evidence that females are better at emotion recognition than males (β = 0.026, 95% CI [0.015, 0.036], $t(911) = 4.82$, $p < .001$) and lower maternal social class was associated with decreased emotion recognition (β = -0.038, 95% CI [-0.058, -0.019], $t(911) = -3.83$, $p < .001$). Increased accuracy on the n-back task was associated with better overall emotion recognition on the Bristol ERT (β = 0.023, 95% CI [0.018, 0.029], $t(911) = 8.53$, $p < .001$). There was no evidence for an association with age (β = 0.001, 95% CI [-0.006, 0.008], $t(911) = 0.32$, $p < .75$).

6.3.3.3 By emotion results

A multiple linear regression was used to investigate the association between adult injury on emotion recognition accuracy (unbiased hit rate) and bias (total number of times an emotion was selected) for each individual emotion.

There was no evidence for differences in emotion recognition accuracy between the three injury groups for any individual emotion (Table 6-4). There was evidence for bias in the mild TBI as participants who reported having a mild TBI were more likely to choose angry to label emotions comparted to both the participants with no injury (β = 0.69, 95% CI [0.22, 1.16], $t(1143) = 2.88$, $p = .004$) and participants with broken bones (β = 0.86, 95% CI [0.18, 1.54], $t(1143) = 2.49$, $p = .013$). An effect size calculation showed that these effects are small, Hedges' $g_s = 0.21$ and 0.26 respectively. The common language effect size indicates that the likelihood of a response bias was only slightly above chance (CL = 0.56 and 0.57). Nevertheless, there was evidence that this association persists, getting weaker after adjusting for social demographic factors, but getting stronger again after adjusting for working memory (Table 6-5).

Given that there was no evidence for increased emotion recognition accuracy for angry faces but increased selection of angry as a response for the mild TBI group the regression analysis was repeated using raw hits for each emotion as the outcome (see Appendix B, section 8.18). The results show that there was a trend towards increased number of hits in the mild TBI group although there was only weak evidence for a difference between the mild TBI group and the other two injury groups.

Table 6-4

Results from regression analysis investigating impact of injury on unbiased hit rate for each emotion

Time of Injury		Unadjusted					Model 1					Model 2				
Age 17 to 25		n	Estimate	95% CI		p	n	Estimate	95% CI		p	n	Estimate	95% CI		p
Emotion	Injury			LL	UL				LL	UL				LL	UL	
Angry	mTBI vs None	1,146	-0.002	-0.005	0.000	0.08	940	-0.002	-0.004	0.001	0.29	831	-0.002	-0.005	0.001	0.19
	BB vs None		-0.002	-0.005	0.001	0.28		0.000	-0.004	0.003	0.88		-0.001	-0.005	0.003	0.69
	mTBI vs BB		-0.001	-0.004	0.003	0.77		-0.001	-0.005	0.003	0.55		-0.001	-0.006	0.003	0.61
Disgust	mTBI vs None		0.002	-0.001	0.004	0.23		0.001	-0.002	0.004	0.37		0.002	-0.001	0.005	0.32
	BB vs None		0.000	-0.004	0.003	0.82		-0.001	-0.005	0.003	0.59		-0.001	-0.005	0.003	0.53
	mTBI vs BB		0.002	-0.002	0.006	0.30		0.002	-0.002	0.006	0.27		0.003	-0.002	0.007	0.23
Fear	mTBI vs None		-0.001	-0.005	0.003	0.56		-0.002	-0.006	0.003	0.49		-0.003	-0.008	0.002	0.25
	BB vs None		-0.005	-0.010	0.000	0.07		-0.003	-0.009	0.002	0.24		-0.004	-0.011	0.002	0.21
	mTBI vs BB		0.004	-0.002	0.010	0.23		0.002	-0.005	0.008	0.59		0.001	-0.006	0.009	0.74
Happy	mTBI vs None		0.000	-0.002	0.002	0.92		0.000	-0.002	0.003	0.84		0.001	-0.002	0.004	0.43
	BB vs None		-0.003	-0.006	0.000	0.05		-0.002	-0.005	0.001	0.23		-0.003	-0.006	0.001	0.19
	mTBI vs BB		0.003	-0.001	0.006	0.10		0.002	-0.002	0.006	0.24		0.004	-0.001	0.008	0.09
Sad	mTBI vs None		0.000	-0.003	0.002	0.83		0.000	-0.003	0.003	0.93		0.000	-0.003	0.003	0.95
	BB vs None		-0.003	-0.006	0.000	0.09		-0.002	-0.005	0.002	0.36		-0.001	-0.005	0.003	0.62
	mTBI vs BB		0.002	-0.001	0.006	0.19		0.002	-0.003	0.006	0.46		0.001	-0.004	0.006	0.69
Surprise	mTBI vs None		-0.002	-0.004	0.000	0.07		-0.002	-0.004	0.001	0.23		-0.002	-0.005	0.001	0.19
	BB vs None		-0.003	-0.005	0.000	0.08		-0.003	-0.006	0.000	0.09		-0.002	-0.005	0.002	0.39
	mTBI vs BB		0.000	-0.003	0.004	0.79		0.001	-0.002	0.005	0.51		0.000	-0.004	0.004	0.91

Note: Sample size reduces for each model as participants who are missing covariate data used in that model are excluded. mTBI refers to mild Traumatic Brain Injury. BB refers to Broken Bones. None refers to No Injury. Estimate is regression coefficient from a multiple linear regression. LL is the lower limit and UL the upper limit of the 95% Confidence Interval. Unadjusted: Impact of receiving an injury on emotion recognition performance measured using the Bristol ERT at age 24. Model 1: As Unadjusted and additionally adjusted for age at 24 clinic, sex at birth, and maternal social class. Model 2: As Model 1 and additionally adjusted for working memory (2-back) at age 24.

Table 6-5

Results from regression analysis investigating impact of injury on number of responses given for each emotion

Time of Injury		Unadjusted					Model 1					Model 2				
Age 17 to 25		n	Estimate	95% CI		p	n	Estimate	95% CI		p	n	Estimate	95% CI		p
Emotion	Injury			LL	UL				LL	UL				LL	UL	
Angry	mTBI vs None	1,146	0.692	0.220	1.164	< .01	940	0.559	0.029	1.089	0.04	831	0.853	0.295	1.412	< .01
	BB vs None		-0.167	-0.759	0.424	0.58		-0.189	-0.861	0.482	0.58		-0.200	-0.945	0.545	0.60
	mTBI vs BB		0.859	0.181	1.537	0.01		0.749	-0.016	1.513	0.05		1.053	0.210	1.897	0.01
Disgust	mTBI vs None		-0.377	-0.951	0.197	0.20		-0.254	-0.906	0.397	0.44		-0.253	-0.949	0.442	0.47
	BB vs None		-0.116	-0.835	0.603	0.75		0.296	-0.529	1.122	0.48		0.094	-0.835	1.022	0.84
	mTBI vs BB		-0.261	-1.086	0.563	0.53		-0.551	-1.491	0.390	0.25		-0.347	-1.397	0.704	0.52
Fear	mTBI vs None		-0.102	-0.725	0.521	0.75		-0.049	-0.768	0.669	0.89		-0.026	-0.806	0.753	0.95
	BB vs None		-0.513	-1.293	0.268	0.20		-0.623	-1.533	0.287	0.18		-0.247	-1.288	0.794	0.64
	mTBI vs BB		0.411	-0.484	1.305	0.37		0.574	-0.462	1.610	0.28		0.221	-0.957	1.398	0.71
Happy	mTBI vs None		-0.146	-0.984	0.692	0.73		-0.072	-1.023	0.879	0.88		-0.350	-1.380	0.680	0.51
	BB vs None		0.816	-0.234	1.866	0.13		0.606	-0.598	1.810	0.32		0.757	-0.618	2.133	0.28
	mTBI vs BB		-0.962	-2.165	0.241	0.12		-0.678	-2.049	0.694	0.33		-1.107	-2.664	0.449	0.16
Sad	mTBI vs None		-0.222	-0.892	0.449	0.52		-0.328	-1.102	0.446	0.41		-0.421	-1.258	0.415	0.32
	BB vs None		0.105	-0.735	0.946	0.81		0.007	-0.974	0.987	0.99		-0.063	-1.179	1.054	0.91
	mTBI vs BB		-0.327	-1.290	0.636	0.51		-0.335	-1.451	0.782	0.56		-0.359	-1.622	0.905	0.58
Surprise	mTBI vs None		0.155	-0.449	0.759	0.61		0.144	-0.545	0.833	0.68		0.198	-0.553	0.948	0.61
	BB vs None		-0.126	-0.882	0.631	0.74		-0.097	-0.969	0.776	0.83		-0.341	-1.343	0.660	0.50
	mTBI vs BB		0.281	-0.587	1.148	0.53		0.241	-0.754	1.235	0.63		0.539	-0.595	1.673	0.35

Note: Sample size reduces for each model as participants who are missing covariate data used in that model are excluded. mild TBI refers to mild Traumatic Brain Injury. BB refers to Broken Bones. None refers to No Injury. Estimate is regression coefficient from a multiple linear regression. LL is the lower limit and UL the upper limit of the 95% Confidence Interval Unadjusted: Impact of receiving an injury on emotion recognition performance measured using the Bristol ERT at age 24. Model 1: As Unadjusted and additionally adjusted for age at 24 clinic, sex at birth, and maternal social class. Model 2: As Model 1 and additionally adjusted for working memory (2-back) at age 24

6.3.3.4 Sensitivity analysis

The analysis for both overall emotion recognition and the by emotion outcomes was repeated with outliers on the Bristol ERT excluded. Outliers were defined as scores more than 1.5 times the interquartile range above the 3rd or below the 1st quartile on total hits. A total of 11 participants were excluded, but there was no change in the results (Appendix C, section 8.19).

6.3.3.5 Step 1 - Results Summary

There was no evidence for a difference in emotion recognition accuracy between the mild TBI group and the no injury group or the broken bones group. The results indicate that there is no association between mild TBI and overall emotion recognition or individual emotion recognition accuracy on the Bristol ERT. Despite there not being a change in the ability to recognise emotions after mild TBI, there was evidence that having a mild TBI was associated with an increased response of angry when identifying emotions. This could be indicative of a negative attribution bias towards identifying ambiguous faces as angry compared to participants who had broken bones or no injury. However, although there was weak evidence for a higher number of angry hits in the mild TBI group, the observed bias did not result in an increase in accuracy for angry faces. The dissociation between accuracy and bias was possibly due to the fact that there was one target emotion and five non-target emotions, which results in 80 possible false alarm options but only 16 possible hits for each emotion. Hits and false alarms do not necessarily mirror each other when there are multiple distractors (Megreya & Burton, 2007).

6.3.4 Step 2 – Association between mild TBI and anxiety at age 24

6.3.4.1 Demographics

A total of 3639 participants competed the CIS-R at age 24 and thus had information about Generalised Anxiety Disorder available for this timepoint. Again, only participants with injury data available were included in this analysis, n = 1161 for injury between age 17 and age 25, and n = 1934 for injury between birth and age 25. Descriptive statistics for these participants are presented in Table 6-6.

Table 6-6

Descriptive statistics for outcome variable (GAD) and confounders by injury group

Injury between ages 17 and 25										
	Total		No Injury		mild TBI		Broken Bones		Injury group comparison	
	n = 1161		n = 753		n = 255		n = 153			
Variable	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**χ^2**	***p***
Female	795	68%	556	74%	152	60%	87	57%	28.9	< .01
Social Class IV & V	63	7%	45	7%	11	5%	7	6%	0.9	0.64
GAD at 24	108	9%	63	8%	36	14%	9	6%	9.9	< .01

Injury between birth and age 25										
	Total		No Injury		mild TBI		Broken Bones		Injury group comparison	
	n = 1934		n = 620		n = 544		n = 770			
Variable	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**&**	**χ^2**	***p***
Female	1208	62%	464	75%	298	55%	42	58%	61	< .01
Social Class IV & V	114	7%	35	7%	33	7%	46	7%	0.16	0.92
GAD at 24	181	9%	54	9%	74	14%	53	7%	17.4	< .01

Note: Sample size can vary due to missing covariate data. Social class variable is maternal social class. GAD is Generalised Anxiety Disorder. Comparisons are chi-squared tests.

There was a higher proportion of females in the adult injury data set compared to the lifetime injury data set, but generally the proportion of participants across variables are comparable between the two data sets. The group comparison indicated that there was a difference in the proportion of females between the no injury, mild TBI, and broken bones groups. Given that the proportion of females was lower in the mild

TBI and broken bones groups compared to the no injury group, this suggests that males were more likely to sustain an injury than females. There was no evidence for a difference in social class across the injury groups. There was evidence for a difference in the proportion of participants with GAD across the injury groups, such that prevalence of GAD was higher in the mild TBI group than the broken bone and no injury groups.

6.3.4.2 Association between injury and GAD

A logistic regression with a binary outcome for the presence of GAD was used to analyse the association between mild TBI and GAD in both the adult injury and lifetime injury data sets (Appendix D. section 8.20). The results were used to calculate a odds ratio to show the likelihood of having GAD in the mild TBI group compared to the no injury group. It was not possible to do this in a meaningful way for the broken bones group, as there were less than 10 individuals in the adult injury broken bones group that were identified as having GAD. Nevertheless, the trend in the data suggests that having a broken bone did not increase the likelihood of having GAD at 24 given the percentage of participants with GAD was lower in that group than the no injury group (Table 6-6). There was evidence that having a mild TBI increased the likelihood of having GAD at 24 in both the adult injury and lifetime injury data sets, and the odds increased when sociodemographic variables were controlled for (Table 6-7). Results indicated that participants who have had an adult mild TBI are almost twice as likely to have GAD at age 24 (an odds ratio of 1 meaning that there was no difference between the groups). However, once the results for the adult injury group were adjusted based on whether participants had GAD at age 17 there was no longer any evidence for an increased likelihood of GAD in the mild TBI group.

6.3.4.3 Step 2 - Results Summary

The results from the second step of the analysis indicate that there was no association between mild TBI and presence of Generalised Anxiety Disorder at 24 in this sample. Despite evidence for an increased likelihood of GAD after mild TBI when only sociodemographic factors are controlled for, it seems that anxiety prior to obtaining the mild TBI was driving the observed association. It is not possible to make causal inferences, but a possible explanation is that participants who are anxious were more likely to report having had a head injury in this sample.

Table 6-7

Likelihood of having GAD at 24 after having a mild TBI compared to the no injury group

Time of Injury	Unadjusted					Model 1					Model 2				
	n	Odds ratio	95% CI		p	n	Odds ratio	95% CI		p	n	Odds ratio	95% CI		p
			LL	UL				LL	UL				LL	UL	
Age 17 to 25	1,161	1.80	1.15	2.77	0.01	951	1.90	1.13	3.18	0.01	778	1.34	0.67	2.57	0.39
Birth to age 25	1,934	1.65	1.14	2.40	0.01	1,593	1.88	1.22	2.92	< .01	1,221	-	-	-	-

Note: Sample size reduces for each model as participants who are missing covariate data used in that model are excluded. A logistic regression with Generalised Anxiety Disorder at 24 as the outcome was conducted and used to calculate odds ratios indicating the likelihood of having GAD in the mild TBI injury group compared to the no injury group. LL is the lower limit and UL the upper limit of the 95% Confidence Interval.

Models:
Unadjusted: Odds of having Generalised Anxiety Disorder at age 24 after mild TBI compared to participants without an injury.
Model 1: As Unadjusted and additionally adjusted for age at 24 clinic, sex at birth, and maternal social class
Model 2: As Model 1 and additionally adjusted for presence of Generalised Anxiety Disorder at age 17 years and 6 months

6.3.5 Step 3 – Association between anxiety and emotion recognition at age 24

6.3.5.1 Demographics

Of the 3551 participants who completed the Bristol ERT at around age 24, 3291 participants had information about Generalised Anxiety Disorder (GAD) available and 3455 indicated whether they had been anxious in the week before completing the Bristol ERT. Descriptive statistics are presented in Table 6-8 and Table 6-9, respectively. Results from the group comparisons indicate that a higher proportion of females have GAD in this sample, but there was no evidence for a difference in social class in the group with GAD compared to no GAD. Similarly, a higher proportion of females reported being anxious in the week before completing the Bristol ERT, but there was no evidence for a difference in social class between the groups. As in step 1 of this analysis, there was evidence for an effect of age across groups in both the GAD and anxiety during the past week, however the mean age was very similar between groups. Further there was evidence that participants with GAD performed worse on the N-back but not on the Bristol ERT compared too controls. Participants with anxiety in the past week performed no different than non-anxious controls on the N-back but better on the Bristol ERT.

Table 6-8

Descriptive statistics for outcome variable and confounders by anxiety group (GAD)

Variable	Total		No GAD		GAD		Anxiety group comparison	
	n = 3291		n = 2955		n = 336			
Categorical	**n**	**%**	**n**	**%**	**n**	**%**	**χ^2**	***p***
Female	2042	62%	1791	61%	251	75%	24.9	< .001
Social Class IV & V	204	8%	177	7%	27	10%	2.4	0.12
Continuous	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	***F-value***	***p***
Age at clinic (years)	24.5	0.8	24.5	0.8	24.6	0.8	6.5	0.01
N-back (d-prime)	2.69	1	2.7	1	2.58	1	4.9	0.03
Bristol ERT (total hits)	66.4	7.9	66.3	7.9	66.6	8	0.4	0.54

Note: Sample size can vary due to missing covariate data. GAD is Generalised Anxiety Disorder Social class variable is maternal social class. Continuous variables were all completed at clinic around 24 years. Comparisons are chi-squared tests for the categorial variables and analysis of variance for the continuous variables.

Table 6-9

Descriptive statistics for outcome variable and confounders by anxiety in past week

Anxiety in the week before completing clinic at age 24								
Variable	Total		No Anxiety		Anxiety		Anxiety group comparison	
	n = 3455		n = 2335		n = 1120			
Categorical	**n**	**%**	**n**	**%**	**n**	**%**	**χ^2**	***p***
Female	2164	63%	1364	58%	800	71%	54.2	< .001
Social Class IV & V	205	7%	146	8%	59	7%	0.57	0.45
Continuous	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	***F-value***	***p***
Age at clinic (years)	24.5	0.8	24.4	0.8	24.5	0.8	8.7	0.003
N-back (d-prime)	2.68	1	2.69	1	2.67	1	0.47	0.49
Bristol ERT (total hits)	66.4	7.9	66	8	67.1	7.4	13.7	< .001

Note: Sample size can vary due to missing covariate data. Anxiety refers to feeling anxious in the week before completing the outcome variable. Social class variable is maternal social class. Continuous variables were all completed at clinic around 24 years. Comparisons are chi-squared tests for the categorial variables and analysis of variance for the continuous variables.

6.3.5.2 Overall emotion recognition accuracy

Multiple linear regression with overall emotion recognition on the Bristol ERT as the outcome shows that there was no evidence for a change in emotion recognition in participants that had GAD compared to those who do not (β = 0.006, 95% CI [-0.015, 0.012], $t(3289) = 0.62$, $p = .54$). There was no change in results when the model was adjusted for sociodemographic variables and working memory performance (Table 6-10). However, there was evidence for improved emotion recognition if participants reported having been anxious in the week before completing the Bristol ERT (β = 0.01, 95% CI [0.005, 0.017], $t(3453) = 3.7$, $p < .001$). Evidence for this association persisted after sociodemographic variables and working memory were controlled for (Table 6-10). However, an effect size calculation showed that the effect was very small, Hedges' g_s = 0.13 and the common language effect size indicated that the likelihood of having improved emotion recognition if anxiety was reported in the week before clinic was only slightly above chance (CL = 0.53).

Table 6-10

Results from regression analysis investigating impact of anxiety on overall emotion recognition

	Unadjusted					Model 1					Model 2				
	n	Estimate	95% CI		p	n	Estimate	95% CI		p	n	Estimate	95% CI		p
			LL	UL				LL	UL				LL	UL	
Generalised Anxiety Disorder	3,291	0.003	-0.006	0.012	0.53	2,633	0.000	-0.010	0.011	0.93	2,588	0.002	-0.008	0.012	0.73
Anxiety in past week	3,455	0.011	0.005	0.017	<0.01	2,754	0.009	0.003	0.016	0.01	2,706	0.009	0.002	0.015	0.01

Note: Sample size reduces for each model as participants who are missing covariate data used in that model are excluded. Estimate is regression coefficient from a multiple linear regression. LL is the lower limit and UL the upper limit of the 95% Confidence Interval.

Models:
Unadjusted: Impact of having anxiety on emotion recognition performance measured using the Bristol ERT at age 24
Model 1: As Unadjusted and additionally adjusted for age at 24 clinic, sex at birth, and maternal social class
Model 2: As Model 1 and additionally adjusted for working memory (n-back) at age 24

Results from the fully adjusted models for both GAD and anxiety in the past week indicated that being female was associated with better emotion recognition performance (β = 0.021, 95% CI [0.016, 0.028], *t*(2582) = 7.02, *p* <.001 and β = 0.021, 95% CI [0.015, 0.027], *t*(2700) = 6.86, *p* <.001). Lower maternal social class was associated with poorer emotion recognition performance (β = -0.019, 95% CI [-0.03, -0.008], *t*(2582) = -3.33, *p* <.001 and β = -0.017, 95% CI [-0.029, -0.007], *t*(2700) = -3.12, *p* <.001). Finally, having better working memory performance was associated with increased emotion recognition performance (β = 0.023, 95% CI [0.02, 0.026], *t*(2582) = 14.7, *p* <.001 and β = 0.021, 95% CI [0.019, 0.025], *t*(2700) = 14.6, *p* <.001).

6.3.5.3 By emotion results

As in step one, a multiple linear regression was used to investigate the association between adult injury on emotion recognition accuracy and responses for each individual emotion. There was no evidence for an association between GAD and unbiased hit rate for any emotion. There was evidence that participants who had been anxious in the week before completing the Bristol ERT had a better unbiased hit rate for fearful faces compared to participants who were not anxious (Table 6-11). Evidence for this association persisted after controlling for sociodemographic factors and working memory. This suggests that the overall increase in emotion recognition could be driven by increased accuracy in identifying fearful faces. However, analysis of the raw hits indicated that participants with anxiety in the past week scored more highly on anger, disgust, fear, and sadness (Appendix E, section 8.21). There is no evidence of an association between GAD or anxiety in the past week on the number of responses given for each emotion (Table 6-12).

6.3.5.4 Step 3 - Results Summary

Results from the third step of this analysis suggest that having GAD at age 24 does not influence emotion recognition performance in this sample, but recent experience of anxiety is associated with increased emotion recognition performance. The by emotion results showed that the improvement in performance was for negative emotions only and when considering the number of misidentifications made for each emotion the recently anxious group had higher recognition accuracy only for fearful faces. However, anxiety in the past week was based on the response to a single item, which may not be a very stable or accurate measure of anxiety.

Table 6-11

Results from regression analysis investigating impact of anxiety on unbiased hit rate for each emotion

		Unadjusted					Model 1					Model 2				
Anxiety	**Emotion**	**n**	**Estimate**	**95% CI**		**p**	**n**	**Estimate**	**95% CI**		**p**	**n**	**Estimate**	**95% CI**		**p**
				LL	**UL**				**LL**	**UL**				**LL**	**UL**	
Generalised Anxiety	Angry	3,291	0.000	-0.002	0.002	0.92	2,633	0.000	-0.002	0.002	0.94	2,588	0.000	-0.002	0.002	0.99
Disorder	Disgust		0.001	-0.001	0.003	0.31		0.000	-0.002	0.003	0.77		0.001	-0.002	0.003	0.61
	Fear		0.000	-0.003	0.004	0.84		0.000	-0.004	0.004	0.91		0.001	-0.003	0.004	0.76
	Happy		0.001	-0.001	0.003	0.36		0.001	-0.001	0.003	0.34		0.001	-0.001	0.003	0.25
	Sad		0.000	-0.003	0.002	0.65		-0.002	-0.004	0.001	0.18		-0.002	-0.004	0.001	0.16
	Surprise		0.000	-0.002	0.002	0.97		-0.001	-0.003	0.001	0.59		0.000	-0.002	0.002	0.74
Anxiety in past week	Angry	3,455	0.001	0.000	0.002	0.11	2,754	0.001	-0.001	0.002	0.20	2,706	0.001	-0.001	0.002	0.21
	Disgust		0.001	0.000	0.003	0.05		0.001	-0.001	0.002	0.40		0.001	-0.001	0.002	0.44
	Fear		0.003	0.001	0.005	<0.01		0.003	0.001	0.006	0.01		0.003	0.001	0.005	0.02
	Happy		0.001	0.000	0.002	0.20		0.001	-0.001	0.002	0.31		0.001	-0.001	0.002	0.27
	Sad		0.001	0.000	0.002	0.11		0.001	0.000	0.003	0.17		0.001	-0.001	0.002	0.22
	Surprise		0.001	0.000	0.002	0.26		0.000	-0.001	0.002	0.48		0.000	-0.001	0.002	0.58

Note: Sample size reduces for each model as participants who are missing covariate data used in that model are excluded. Estimate is regression coefficient from a multiple linear regression. LL is the lower limit and UL the upper limit of the 95% Confidence Interval.
Unadjusted: Impact of having anxiety on emotion recognition performance measured using the Bristol ERT at age 24. Model 1: As Unadjusted and additionally adjusted for age at 24 clinic, sex at birth, and maternal social class. Model 2: As Model 1 and additionally adjusted for working memory (n-back) at age 24

Table 6-12

Results from regression analysis investigating impact of injury on number of responses for each emotion

		Unadjusted					Model 1					Model 2				
Anxiety	Emotion	n	Estimate	95% CI		p	n	Estimate	95% CI		p	n	Estimate	95% CI		p
				LL	UL				LL	UL				LL	UL	
Generalised Anxiety Disorder	Angry	3,291	0.132	-0.247	0.511	0.49	2,633	0.048	-0.386	0.481	0.83	2,588	0.053	-0.383	0.488	0.81
	Disgust		0.007	-0.458	0.473	0.98		0.105	-0.427	0.637	0.70		0.072	-0.462	0.606	0.79
	Fear		0.120	-0.397	0.637	0.65		-0.041	-0.628	0.545	0.89		0.024	-0.565	0.614	0.94
	Happy		-0.567	-1.250	0.116	0.10		-0.655	-1.432	0.121	0.10		-0.700	-1.481	0.082	0.08
	Sad		0.317	-0.219	0.854	0.25		0.560	-0.049	1.169	0.07		0.594	-0.020	1.207	0.06
	Surprise		-0.010	-0.495	0.474	0.97		-0.016	-0.565	0.533	0.95		-0.043	-0.591	0.505	0.88
Anxiety in past week	Angry	3,455	0.152	-0.086	0.391	0.21	2,754	0.165	-0.109	0.438	0.24	2,706	0.124	-0.152	0.400	0.38
	Disgust		0.160	-0.133	0.453	0.28		0.258	-0.077	0.592	0.13		0.257	-0.080	0.594	0.14
	Fear		0.040	-0.287	0.367	0.81		-0.064	-0.435	0.307	0.74		-0.061	-0.435	0.313	0.75
	Happy		-0.286	-0.716	0.144	0.19		-0.294	-0.785	0.196	0.24		-0.311	-0.806	0.184	0.22
	Sad		0.099	-0.239	0.438	0.56		0.161	-0.223	0.546	0.41		0.178	-0.210	0.567	0.37
	Surprise		-0.166	-0.472	0.141	0.29		-0.226	-0.574	0.123	0.20		-0.187	-0.535	0.161	0.29

Note: Sample size reduces for each model as participants who are missing covariate data used in that model are excluded. Estimate is regression coefficient from a multiple linear regression. LL is the lower limit and UL the upper limit of the 95% Confidence Interval.
Unadjusted: Impact of having anxiety on emotion recognition performance measured using the Bristol ERT at age 24. Model 1: As Unadjusted and additionally adjusted for age at 24 clinic, sex at birth, and maternal social class. Model 2: As Model 1 and additionally adjusted for working memory (n-back) at age 24

6.4 Discussion

The aim of this study was to investigate associations between emotion recognition, mild TBI, and anxiety in the ALSPAC cohort. There was no evidence that having mild TBI, or GAD were associated with changes in emotion recognition accuracy in ALSPAC. There was weak evidence that having anxiety in the week before completing the Bristol ERT was associated with increased emotion recognition accuracy, which seems to be specific to fearful faces. There was also weak evidence that mild TBI was associated with a response bias towards angry faces in this sample, but neither measure of anxiety was associated with a response bias on the Bristol ERT. The lack of overlap between these results means it is unlikely that anxiety was impacting changes in emotion recognition after mild TBI in this cohort. This is further supported by the fact that although there was evidence for an increased likelihood of having GAD after adult mild TBI, this association was fully explained when the analysis was adjusted for GAD prior to the injury. Given the lack of evidence for associations between mild TBI and anxiety the results for each step of the analysis are discussed separately, and mediation analyses were not conducted.

6.4.1 Step 1 - Association between mild TBI and emotion recognition

Contrary to our prediction, there was no evidence for changes in emotion recognition accuracy after mild TBI in this sample. These findings are in line with other studies investigating emotion recognition after mild TBI, that have not reported changes in emotion recognition after mild TBI measured using other emotion recognition tasks (McLellan & McKinlay, 2013; Theadom et al., 2019). Theadom and colleagues (2019) did report a decrease in emotion recognition accuracy for happy stimuli, which we did not find evidence for in the current study. A null result is difficult to interpret, but possible explanations for this difference are that the EET from TASIT is picking up a difficulty in reading non-verbal emotional cues other than facial expressions. Another consideration is that participants with mild TBI were identified differently in the two studies. Given that in ALSPAC the only information available was that participants had a head injury with loss of consciousness, it is possible that the injuries included were milder than those included in the study by Theadom and colleagues (2019).

Analysis of the number of responses made for each emotion showed that there was weak evidence for a response bias towards angry by participants with mild TBI compared to both control groups. Interestingly this did not correspond to an increased

accuracy for angry stimuli, which is likely due to the number of highly ambiguous stimuli included in the Bristol ERT. The results are indicative of an attribution bias toward identifying ambiguous stimuli as angry. Similarly, young offenders with mild TBI have been shown to have a bias towards perceiving faces as angry (Penton-Voak et al., 2013). Given that hostile attribution bias has been linked to aggression after TBI (D. Neumann et al., 2021), it is possible that response bias observed in participants with adult mild TBI in ALSPAC is related to an increase in aggression after injury. It was not possible to investigate this as part of this study, but notably Kennedy and colleagues (2017) found that participants with childhood mild TBI in ALSPAC were at higher risk of having conduct problems, so it would be worth investigating further.

6.4.2 Step 2 - Association between mild TBI and anxiety

The main hypothesis was that mild TBI would be associated with higher likelihood of having GAD. There was evidence of that association in this cohort, but results suggest that the observed association was driven by preinjury GAD and not directly related to having a mild TBI. There is evidence that in moderate to severe TBI, anxiety prior to TBI is associated with an increased risk of post injury anxiety (Gould et al., 2011). The results in this study could indicate that the same is true for mild TBI but it is also possible that participants with anxiety were more likely to report a head injury when completing the ALSPAC questionnaires. A limitation of the current analysis is that there is potential overlap between the diagnosis of GAD at 17 years and 6 months, and the timeframe of injury between 17 and 25. Due to reduction in sample size when only reports of mild TBI between 18 and 25 years were used it was not possible to remove that overlap from this analysis. Another consideration is that anxiety after moderate to severe TBI is highest in the first year post injury (Alway et al., 2016; Gould et al., 2011). Given that the timeframe of injury was up to 6 years in this study, it is possible that short-term increases in anxiety after mild TBI were not captured and that preinjury anxiety is predictive of long-term anxiety after mild TBI. Alternatively, GAD at age 17 simply predicts GAD at age 24. This study only considered presence of GAD after mild TBI, but there is evidence that mild TBI is associated with increased levels of anxiety without diagnosis of GAD (Gould et al., 2011; Osborn et al., 2016). This means there may be increased anxiety after mild TBI that is not captured by using GAD as the outcome measure for anxiety. A detailed analysis all different anxiety disorders and the associations between pre and post injury anxiety disorders was not possible given the variables available in ALSPAC at this time. There was also evidence for an increased

likelihood of having a diagnosis of GAD after mild TBI at any age (between birth and 25 years). This indicates that there is a link between mild TBI and GAD in this cohort and suggests that it is not just specific to mild TBI in adulthood. Further analysis is needed to better understand the observed results and investigate the directionality of this association. In sum, the results are in line with the current understanding that mild TBI is associated with worsened anxiety (Osborn et al., 2016; Ponsford et al., 2018). It was not possible to draw conclusion about the causal nature of this association in this study.

6.4.3 Step 3 - Association between anxiety and emotion recognition

The main hypothesis for this part of the analysis was that both GAD and recent anxiety would be associated with decreased emotion recognition on the Bristol ERT. Contrary to our prediction, GAD was not associated with emotion recognition accuracy and recent anxiety was associated with an increase, not a decrease, in emotion recognition accuracy. It is possible that the question about feeling anxious nervous or tense in the past week is tapping into the construct of trait anxiety and not state anxiety given that there is evidence for trait anxiety being positively correlated to emotion recognition accuracy (Attwood et al., 2017; Surcinelli et al., 2006). Although number of hits increased for all negative emotions in the recent anxiety group, only fearful faces were recognised more accurately, which is congruent with findings in the study by Surcinelli and colleagues (2006). However, the findings in this study should be interpreted with caution because the recent anxiety variable was based on a single question and the size of the effect observed was very small. Even if there is a true effect there may not be any meaningful functional differences. Another consideration is that there has been mixed evidence regarding the association between trait anxiety and emotion recognition, with several studies not reporting any association (Cooper et al., 2008; Suslow et al., 2019). Many studies investigating links between trait anxiety and emotion recognition have used the State-Trait Anxiety Inventory (Spielberger, 1983) to measure trait anxiety. There is increasing evidence that the trait version of the State-Trait Anxiety Inventory is strongly correlated with both depression and anxiety and should be considered a measure of negative affect, rather than trait anxiety (Knowles & Olatunji, 2020). Due to high comorbidity between depression and anxiety in ALSPAC, it was not possible to exclude participants with depression from the current study, so the increased emotion recognition accuracy observed may not be linked to anxiety specifically but related to more general negative affect.

6.4.4 Covariates

In terms of the covariates included in this study the results indicate that being female, and having higher social class were associated with increased emotion recognition on the Bristol ERT. The direction of these associations is congruent with current literature. Females have been shown to have better emotion recognition accuracy than males, especially when lower intensity stimuli are included in the emotion recognition tasks (Hoffmann et al., 2010; Thompson & Voyer, 2014). Having higher social class has been linked to higher scores on emotion recognition tasks, which may be linked to differences in education (Deveney et al., 2018). It is not surprising that there were no age-related associations with emotion recognition in this sample given the limited age range of participants. Incidence of TBI is reported to be higher for males than females (Headway, 2018; Maas et al., 2008). Reflective of this, there was a higher prevalence of males in the mild TBI compared to the non-injury group in the ALSPAC cohort. Conversely, females are more likely to struggle with anxiety than males (McLean et al., 2011), which corresponds with the higher prevalence of females in both the GAD and recent report of anxiety groups in this study.

Working memory performance measured at the same time as emotion recognition was associated with performance on the Bristol ERT, which is understandable given the demands of the task. There was no evidence of a difference in performance between the mild TBI group and the control groups in this study and controlling for working memory did not change associations observed between mild TBI and emotion recognition. Although it is important to consider the impact of working memory or general cognitive ability on Bristol ERT performance the findings are congruent with the current literature that cognitive ability does not account for differences in emotion recognition after TBI (Rosenberg et al., 2015). Notably, Theadom and colleagues (2019) found evidence for an association between cognitive ability and social cognition in their study, but it did not account for the deficits in social cognition they observed after mild TBI. Having GAD was associated with lower working memory performance, which is congruent with research showing that worry linked to GAD is associated with decreased working memory performance (Abushalbaq et al., 2021; Bredemeier & Berenbaum, 2013; Held et al., 2020). Interestingly, Abushalbaq and colleagues (2021) found that GAD was associated with decreased performance on the N-back task but not some of the other working memory tasks they included.

6.4.5 Strengths and Limitations

An advantage of investigating associations between emotion recognition and mild TBI in ALSPAC was that both uninjured and orthopaedic controls were available as control groups. Evidence for a negative attribution bias in the mild TBI group compared to both control groups suggests that change is not simply related to distress associated with having an injury (Curran et al., 2000; Kennedy et al., 2017). It also reduces the probability that the finding is a false positive, which means that negative attribution bias after mild TBI warrants further investigation. An important consideration is the classification of mild TBI used in this study. There is evidence of differences in outcome after mild TBI based on the length of post traumatic amnesia (Dahm & Ponsford, 2015) and arguably post traumatic amnesia is a better indicator of mild TBI than loss of consciousness (Gasquoine, 2020). The information about TBI available in ALSPAC was limited and the mild TBI variable is based on a singular question about loss of consciousness after head injury. There was no information available about length of loss of consciousness, post traumatic amnesia, or time since injury and it is possible that there was recall bias considering participants were asked to report head injuries over a seven-year period. Studies with more detailed information about TBI, such as the cross-sectional study presented in Chapter 5, are needed to develop our understanding of emotion recognition after mild TBI.

Unlike Theadom and colleagues (2019) the current study was able to consider how anxiety at the time of testing may be impacting emotion recognition. A limitation was that the recent anxiety variable was based on a single item and not a validated questionnaire. Furthermore, anxiety and depression are highly comorbid in ALSPAC, so it was not possible to consider the impact of anxiety independent of depression. This limits the interpretation of the findings reported in this study about the associations between emotion recognition and anxiety. Another limitation of this study was that the variables available in ALSPAC were not included specifically to investigate associations between emotion recognition, TBI, and anxiety. This meant that the timing of data collection could have impacted our findings and variable selection may be subject to bias in a different way to other studies in the thesis. For example, participants were asked about mild TBI at around age 25 but completed the emotion recognition task at around age 24, which means it is possible that some of the participants in the mild TBI group did not yet have a head injury when they completed the Bristol ERT. Similarly, participants completed the anxiety measures at age 17 and a half and again at age 24,

but the mild TBI variable included any injury between 17 and 25. Given the large sample size, it is unlikely that these overlaps will have substantially impacted the results, but it is nevertheless possible. Another benefit of the large sample size is that it was possible to detect small effects. It is worth pointing out that the results in this study were not adjusted for the number of comparisons and should be interpreted by looking at strength of evidence instead of dichotomising based on p-values (Lakens et al., 2018; Sterne & Smith, 2001). Assuming that the effects are true, it is important to remember that small effects may not be meaningful in the real world. The ALSPAC sample is not necessarily representative of the wider population as it consists mainly participants with white ethnicity and participants who have experienced adversity are more likely to have dropped out of the study (Fraser et al., 2013; Munafo et al., 2018). Consequently, findings may not generalise to wider populations.

6.4.6 Implications and future directions

This study added to the literature suggesting that emotion recognition may not be impaired after mild TBI in the same way that it is after moderate to severe TBI. Studies have not considered severity of mild TBI when investigating emotion recognition in this population, and future research should consider how mild TBI is defined and whether differences in severity of injury could be impacting findings. The results in this study also indicated that mild TBI could be associated with a negative attribution bias towards angry faces. This is also worth exploring further, especially given the recent findings by Neumann, Sander, and colleagues (2021) about negative attribution bias after TBI and how it is associated with aggression. The fact that GAD did not seem to impact emotion recognition, indicates that it is not contributing to the negative attribution bias observed after mild TBI. Further research is needed to confirm this, as limitations regarding the anxiety variables available and the time at which they were collected mean that it is difficult to make generalisable inferences based on this study alone.

Chapter 7 General Discussion

7.1 Review of aims and summary of main findings

The three main aims of this thesis were outlined in the introductory chapter. The first was to investigate the psychometric properties of the Bristol Emotion Recognition Task (ERT), to help assess whether the task should be used to measure emotion recognition in a clinical setting. The second was to develop our understanding of changes in emotion recognition after both mild and moderate to severe TBI using the Bristol ERT. The third aim was to evaluate whether changes in emotion recognition after TBI could be associated with anxiety. The main findings from each study included in the thesis are summarised in Table 7-1 and then each of the above aims is reviewed in turn.

Table 7-1

Summary of main aim and findings for each chapter in this thesis

Chapter	Main aim	Main findings
Chapter 2	Investigate the test-retest reliability of the Bristol ERT in ALSPAC	The test-retest reliability coefficients indicated that constancy in measurement over time was moderate/marginal (ICC2.1 = 0.56, ICC3.1 = 0.6) Participants showed improved performance at the second session indicative of a practice effect on the Bristol ERT Performance accuracy for individual emotion measured using the unbiased hit rate was associated with poor test-retest reliability
Chapter 3	Construct validation of the Bristol ERT in a neurologically healthy population	Performance on the Bristol ERT was correlated with performance on two well developed measures of emotion recognition (Dynamic ERT & EET from TASIT-S) Evidence that Bristol ERT was correlated with performance on a face discrimination task (GFMT) but not a working memory task (Digit Span Backwards) No evidence that performance on the Bristol ERT was correlated with scores on an alexithymia scale (TAS) Overall, the pattern of results indicated that the Bristol ERT is appropriate as a measure of emotion recognition Split half reliability estimate indicated marginal reliability measurements made using the Bristol ERT (r_{sb} = 0.65) in this study, but Cronbach's alpha was within the adequate range (α = 0.77)

Chapter 4	Use the Bristol ERT to investigate emotion recognition after moderate to severe TBI in a clinical setting	Evidence that people with moderate to severe TBI have decreased overall emotion recognition performance on the Bristol ERT and the size of the effect detected (Hedge's g_s = 1.28) was congruent with a comparable meta-analytic effect size (Hedge's g = 1.1) There was no association between the anxiety subscale of the DASS and overall emotion recognition on the Bristol ERT suggesting that levels of anxiety do not explain the deficit in emotion recognition associated with moderate to severe TBI Both the split half reliability estimate (r_{sb} = 0.8) and Cronbach's alpha (α = 0.86) were indicative of high reliability of measurement for the Bristol ERT in this study
Chapter 5	Use the Bristol ERT to investigate emotion recognition after mild TBI in a community based population	No evidence that overall emotion recognition on the Bristol ERT was associated with mild TBI after age and face discrimination ability (GFMT) were controlled for There was no evidence of an association between anxiety (DASS subscale) and emotion recognition outcomes on the Bristol ERT Increased age and lower scores on the face discrimination task (GFMT) were associated with lower scores on overall emotion recognition on the Bristol ERT The split half reliability coefficient suggested adequate reliability of measurement on the Bristol ERT (r_{sb} = 0.68), whilst Cronbach's alpha indicated high internal consistency in this study (α = 0.8)
Chapter 6	Investigate associations between emotion recognition, mild TBI, and anxiety in a longitudinal birth cohort (ALSPAC)	No evidence that overall emotion recognition on the Bristol ERT was associated with mild TBI Evidence that having a mild TBI may be associated with a negative attribution bias towards perceiving faces as angry on the Bristol ERT Increased odds of having GAD in the mild TBI group, but this was explained by presence of GAD prior to injury No evidence that having GAD was associated with performance on the Bristol ERT, but reports of having experienced anxiety in the week before completing the Bristol ERT was associated with an increase in overall emotion recognition accuracy

Note: ALSPAC is the Avon Longitudinal Study of Parents and Children prospective cohort. DASS is Depression, Anxiety, and Stress Scale. EET from TASIT-S is the Emotion Evaluation Test from the short version of The Awareness of Social Inferences Test. ERT is Emotion Recognition Task. GAD is Generalised Anxiety Disorder. GFMT is the Glasgow Face Matching Test. ICC is the Intraclass Correlation Coefficient. TAS is Toronto Alexithymia scale. TBI is Traumatic Brain Injury.

7.2 Aim One - Psychometric properties of the Bristol ERT

The first aim of this thesis was to develop our understanding of the psychometric properties of the Bristol ERT. This is important because without information about reliability and validity it is not possible to effectively draw inferences about emotion recognition based on the Bristol ERT outcome measures (Bowden, 2017). Reliability coefficients are used to evaluate whether measurements made using the Bristol ERT are consistent given stable conditions (Streiner et al., 2015). Assessing validity is an continual process of evaluating whether valid inferences about emotion recognition can be made based on the Bristol ERT outcomes (Streiner et al., 2015).

7.2.1 Reliability

Two different types of reliability for measurement of overall emotion recognition (total hits) using the Bristol ERT were assessed. Test-retest reliability is an indicator of the stability of measurement over time, whilst internal consistency reliability assesses the extent to which items within a task are measuring the same construct. Test-retest reliability in ALSPAC (Chapter 2) was marginal at best, whilst internal reliability estimates were adequate to high across the studies presented in Chapters 3, 4, and 5. These reliability estimates need to be considered in conjunction when evaluating whether to use the Bristol ERT to assess emotion recognition in a clinical setting (Sherman et al., 2011). Low test-retest reliability estimates means that it might not be appropriate to use the Bristol ERT to evaluate change in individual performance over time. However, internal consistency estimates suggest that performance on the Bristol ERT could provide a useful indication of whether an individual is experiencing emotion recognition difficulties post TBI. Especially as reliability estimates for measurements in the moderate to severe TBI group were comparable to reliability estimates for neurologically healthy controls (TBI group r_{sb} = 0.7, and control group r_{sb} = 0.74).

Notably, there was still evidence of considerable individual variation in performance for the number of total hits on the Bristol ERT. The Smallest Real Difference calculated in Chapters 2 to 5 was consistently around 13 hits. This could provide useful information on individual performance if scores could be compared to age specific normative data. However, it is a big difference in scores considering there are only 96 trials on the version of the Bristol ERT used in this thesis. The reliability estimates also clearly show that overall emotion recognition should not be used for diagnosis or clinical decision making as there was too much variation in individual

performance (Sherman et al., 2011; Strauss et al., 2006). The reliability estimates calculated in Chapter 3 for the Dynamic ERT (Montagne et al., 2007) and reliability estimates for the EET from TASIT (McDonald et al., 2006) are better than the estimates associated with the version of the Bristol ERT used in this thesis. This means these tasks would likely be more suitable to assess emotion recognition unless adjustments to the Bristol ERT can be made to improve the reliability measurement based on total number of hits. See section 7.2.3 for a discussion of changes that could be made to develop the Bristol ERT as a measure of emotion recognition for individual assessment.

Reliability estimates broadly indicate that the Bristol ERT is appropriate for research purposes (Ponterotto & Ruckdeschel, 2007), although using it to assess individual change in performance is not recommended. Hedge and colleagues (2018) have argued that when considering group differences in a research context, a task is considered reliable if it consistently replicates a given effect with a consistent effect size. The Bristol ERT was successfully used to replicate deficits in emotion recognition associated with moderate to severe TBI (Chapter 4) so can be considered reliable in that sense. It is also worth considering that reliability estimates in the moderate to severe TBI study were higher than in the mild TBI (Chapter 5) and no TBI validation (Chapter 3) studies. Recruitment for the latter two studies was wholly online, whilst participants with moderate to severe TBI all attended supervised testing session (although control participants did not). It is possible that online recruitment impacted quality of data and this resulted in lower reliability estimates. Equally, reliability in those samples could simply have been lower, which highlights the importance of reporting reliability estimates for all studies (Parsons et al., 2019). Cronbach's alpha was higher than the split half reliability estimates in all the studies where it was possible to calculate internal consistency. There was a small increase in the split half reliability estimates when low intensity trials were removed, whilst Cronbach's alpha tended to decrease slightly. It seems like the method of calculation for these two estimates means they are impacted by different things. The reduction in Cronbach's alpha is probably due to a decrease in the number of trials included (Tavakol & Dennick, 2011), whilst the increase in split half reliability is likely the result of less variation when low intensity trials are removed. This shows that it is important to consider both these factors when adjusting the Bristol ERT. A general consideration is that emotion recognition may be associated with high measurement variability because of state related fluctuations, for example when people are in a state of anxiety (Attwood et al., 2017; Dyer et al., 2022).

7.2.1.1 Emotion specific recognition accuracy

The findings in this thesis indicate that the unbiased hit rate proposed by Wagner (1993) is not a useful measure of individual emotion recognition accuracy for the Bristol ERT. The test-retest reliability coefficients calculated in ALSAPC were low across the board (Chapter 2). As discussed in Chapter 4 it was not possible to make meaningful inferences about emotion specific accuracy after moderate to severe TBI based on the unbiased hit rate. Other tasks, such as TASIT (McDonald et al., 2006) and the Dynamic ERT introduced in Chapter 3 (Kessels et al., 2014; Montagne et al., 2007) use the number of hits for each emotion as an emotion specific outcome. Whilst this approach is simple it does not truly capture emotion recognition accuracy, which is why it was not used as on outcome measure in this thesis. For example, a participant could score highly on anger simply because they select anger most of the time, but it does not follow that they can accurately identify anger (Wagner, 1993). An alternative approach to measuring emotion recognition accuracy is using Signal Detection Theory. A sensitivity index can be calculated to assess the ability to discriminate between a target emotion (i.e. signal) and noise, in this case all other emotions included (Stanislaw & Todorov, 1999). As outlined by Wagner (1993) an issue with this approach is that one of the underlying assumptions is that all of the alternative (non-target) emotions are equally likely to be chosen. Furthermore, on a six Alternate Force Choice tasks such as the Bristol ERT there are five times as many distractor trials then target trials. Nevertheless, as demonstrated by Eastwood and colleagues (2020) this approach can be used to evaluate individual emotion recognition outcomes on the Bristol ERT. Further studies are needed to evaluate whether this sensitivity index is an appropriate outcome measure to use in a clinical setting. It could also be worth exploring whether latent growth curve modelling is a better method of accessing individual emotion recognition accuracy given increased test-retest reliability estimates reported by Cecilione and colleagues (2017) when using that approach.

7.2.2 Validity

Having a reliable outcome measure is the basis for valid inferences about emotion recognition to be made using the Bristol ERT. Consequently, understanding and where possible improving reliability estimates for measurements made using the Bristol ERT is important. The internal consistency reliability coefficients for studies presented in this thesis were considered sufficient for emotion recognition inferences to me made

about overall emotion recognition but not emotion specific outcomes. Furthermore, the fact that it was possible to discriminate between participants with moderate to severe TBI and neurologically healthy controls using the Bristol ERT indicates that it is a suitable measure to use in a TBI population.

The construct validation study presented in Chapter 3 indicated that the Bristol ERT is measuring the same construct as the Dynamic ERT developed by Montagne and colleagues (2007). This suggests that the Bristol ERT could be an equivalent task if reliability of measurement can be improved. Khosdelazad and colleagues (2020) recently compared performance on the Facial Expressions of Emotion: Stimuli and Test (FEEST; Young et al., 2002), the Dynamic ERT and the short version of the EET from TASIT (Westerhof-Evers et al., 2014). The correlation between the three tasks was comparable to the results in this thesis, although the correlation between Dynamic ERT and Bristol ERT was slightly higher than the correlation between the FEEST and Dynamic ERT. Khosdelazad and colleagues (2020) suggested that differences in stimuli and difficulty could be underlying their results. It is possible that the Bristol ERT is measuring a construct more closely matched to the Dynamic ERT than other static emotion recognition tasks. Further work is needed to understand how the two facial emotion recognition tasks relate to the full version of the EET from TASIT (McDonald et al., 2006).

Further work is also needed to understand associations between performance on the Bristol ERT and other factors. The Bristol ERT is clearly associated with the ability to discriminate between faces, although results in this thesis show that face perception does not fully account for performance on the Bristol ERT. Contrary to predictions, alexithymia scores were not associated with performance on the Bristol ERT in any of the studies included in this thesis. This supports the argument that overall emotion recognition accuracy on forced choice emotion recognition tasks is not directly linked to one's ability to identify and process emotions (Ihme, Sacher, Lichev, Rosenberg, Kugel, Rufer, Grabe, Pampel, Lepsien, Kersting, Villringer, & Suslow, 2014; Rosenberg et al., 2019). There is evidence that the association between emotion recognition and alexithymia is specific to angry expressions (Rosenberg et al., 2020), which would be worth exploring in the future. In terms of working memory, Bristol ERT scores were not correlated with performance on the Digit Span Backwards task (Chapter 3) but were correlated with performance on the N-back task (Chapter 6). This suggests these two tasks are tapping into different aspects of cognitive ability (Hilbert et al., 2015; Kirchner, 1958; Soveri et al., 2018; St Clair-Thompson, 2010). Differences in association between

digit span tasks and the N-back task have also been observed for Generalised Anxiety Disorder (Abushalbaq et al., 2021), so this pattern of results is not something that is specific to the Bristol ERT or even emotion recognition. A clear limitation of this thesis is that it was not possible to evaluate the impact of processing speed on Bristol ERT performance. Short presentation time of stimuli on the Bristol ERT could mean that performance on the Bristol ERT is particularly sensitive to changes in processing speed. Severity of injury seem to be predictive of difficulties with processing speed in moderate to severe TBI (Svingos et al., 2019). Given that results in this thesis suggests that severity of injury is also associated to performance on the Bristol ERT it is important to assess potential interactions before the task is used for neuropsychological assessment. Notably, Osborne-Crowley and colleagues (2019) have found evidence that contextual processing of emotional facial expressions is preserved even if processing speed is impaired.

In sum, task characteristics clearly play an important part in the measurement of emotion recognition. It is crucial that characteristics of the Bristol ERT and how they could be impacting performance is better understood. Further, to help interpretation, the tasks used to support interpretation of scores should be consistent, otherwise it will be difficult to draw valid inferences. It is also important to remember that validation of the Bristol ERT is an ongoing process and validity of inferences made using the Bristol ERT should be continually evaluated, especially if the task is developed to address some of the issues identified in this thesis.

7.2.3 Potential task development

7.2.3.1 Removing low intensity trials on the Bristol ERT

Reliability estimates associated with measurement on the Bristol ERT could be improved by removing low intensity trials for this task. Decreasing random variance that seems to be associated with those highly ambiguous trials would likely also help improve the emotion specific outcomes. It would be advisable to add more trials with greater than 40% intensity, as simply reducing the number of trials will also impact reliability of measurements made using the Bristol ERT (Tavakol & Dennick, 2011). The morph sequences used to create the stimuli for the Bristol ERT consist of 15 images in total, of which only every second (eight images) was used for the Bristol ERT in this thesis. A better approach could be to use the ten highest intensity images, which correspond to the cut-off point used to create the reduced Bristol ERT for this thesis.

7.2.3.2 Using only four emotions

Another consideration is whether it is useful to include all six emotions on the Bristol ERT. Jack and colleagues (2014) have suggested that dynamic facial expressions are processed sequentially in the brain, starting with basic information and progressing to more complex emotional signals. Given that the stimuli are only presented for a short period of time on the Bristol ERT it is possible that performance on the task is more dependent on the early part of that processing system. Arguably there are only four basic emotion categories at that stage (Jack et al., 2014; Jack et al., 2016), namely happiness, sadness, fear/surprise (potential indicator of approaching danger), and anger/disgust (potential indicator of proximal danger). Potentially only using four stimuli would improve reliability of measurement and validity of inferences made using the Bristol ERT. Further, avoiding confusion between fear/surprise and anger/disgust and could give more accurate measures of individual emotion recognition accuracy. The version of the Bristol ERT used as part of the EMOTICOM battery included only happy, sad, fearful, and angry stimuli for example (Bland et al., 2016). It would be worth exploring whether using the stimuli for surprise as opposed to fear is more appropriate, as fear stimuli are generally considered to be the most difficult to recognise (Rosenberg et al., 2014). Reducing the number of emotions included would obviously also reduce the number of trials. A potential solution is to add another set of stimuli, which could be of a different ethnicity considering the evidence that the four emotion categories could be universally recognised (Jack et al., 2016). Once four emotions are selected it would be good to consider what outcome measure would be most appropriate for individual emotion recognition accuracy. Based on the test-retest reliability data latent growth modelling seems like a promising approach (Cecilione et al., 2017) and intercept scores should be explored as an outcome measure for the Bristol ERT.

7.2.3.3 Combining Bristol ERT with other tasks for a more comprehensive test battery

There is no one right way to measure emotion recognition. The Bristol ERT shows promise as a measure in clinical practice, but it provides information about one part of cognition. It may be worth adding other tasks to capture different aspects of social cognition or help with the interpretation of scores on the Bristol ERT (Castro et al., 2016). Removing low intensity trials that have high ambiguity means that the responses on the Bristol ERT will be less likely to detect attribution bias, so the Bristol EBT tasks could be included as measures of bias. Furthermore, The Cambridge Automated

Neuropsychological Test Battery (CANTAB) owned by Cambridge Cognition Ltd includes tests of processing speed and working memory. Given the recommendations made by Henry and colleagues (2015) these could easily be included as part of a wider assessment if there are concerns about the impact of general cognitive ability on emotion recognition for a particular individual. Another possibility would be to include other measures of social cognition to develop a more rounded assessment tool like the EMOTICOM battery (Bland et al., 2016).

7.3 Aim Two - Emotion recognition on the Bristol ERT and Traumatic Brain Injury

The second aim of this thesis was to use the Bristol ERT to assess emotion recognition after TBI. This included trying to replicate the deficit in emotion recognition associated with moderate to severe TBI (Chapter 4) and investigating whether there is evidence for an association between mild TBI and emotion recognition (Chapter 5 and Chapter 6).

7.3.1 Moderate to Severe Traumatic Brain Injury

A diagnosis of moderate to severe TBI is given when there is no doubt that there has been damage to the brain as the result of an external force (Malec et al., 2007; Menon et al., 2010). For the study presented in Chapter 4 participants with a clinical diagnosis of moderate to severe TBI were recruited to complete the Bristol ERT. There was strong evidence that participants with TBI performed worse on the Bristol ERT than controls, which corresponds with the current literature indicating that there are deficits in emotion recognition after moderate to severe TBI (Babbage et al., 2011; Murphy et al., 2021). Further, the confounders included (face perception, alexithymia, mood, and aggression) were not able to explain the decrease in emotion recognition observed after TBI in our sample. This research shows that the Bristol ERT can be used to assess emotion recognition after TBI and can be used to develop our understanding of difficulties observed in this population. Given the issues around interpreting emotion specific outcomes used in this thesis further work in needed before the Bristol ERT can be used to make inferences about emotion specific accuracy in this population. Although notably, there was no evidence that having moderate to severe was associated with a negative attribution bias on the Bristol ERT or EBT tasks. This suggests that if a bias is observed after TBI it is likely mediated by other factors, for example aggression (Deveney et al., 2019; Neumann et al., 2017; D. Neumann et al., 2021).

7.3.2 Mild Traumatic Brain Injury

The diagnosis of mild TBI is based on lack of evidence for moderate to severe TBI and likelihood that a head injury has resulted in damage to the brain (Gasquoine, 2020; Malec et al., 2007). Both studies investigating emotion recognition after mild TBI in thesis did not find any evidence to suggest a decrease in emotion recognition accuracy in people with mild TBI compared to controls. Although each of these studies individually has limitations the fact that both the cross-sectional study and the cohort study did not find evidence of an association increases the confidence in the results. In context of the two other studies that have investigated emotion recognition after mild TBI (McLellan & McKinlay, 2013; Theadom et al., 2019), there is growing evidence that mild TBI is not associated with decreased emotion recognition in the way that moderate to severe TBI is. However, interpreting lack of evidence (null finding) is difficult as there are often many factors that could lead to a null finding and it is not appropriate to infer a lack of effect based on null findings (Lakens et al., 2020). Instead of trying to find evidence for an association between mild TBI and emotion recognition future studies should investigate whether there is an absence of this association using Bayesian statistics or equivalence testing (Lakens et al., 2020).

In the ALSPAC study (Chapter 6) there was evidence that participants with mild TBI were more likely to label faces as angry, which is indicative of a negative attribution bias. There was no evidence of such a bias in the cross-sectional mild TBI study (Chapter 5). One potential explanation is that due to the larger sample size in the cohort study it was possible to detect a much smaller effect. Given that negative attribution bias seems to be mediated by levels of aggression (D. Neumann et al., 2021) in a TBI population, another explanation is that levels of aggression for the mild TBI group compared to controls were higher in the cohort study. Investigating measures of aggression in ALSPAC was beyond the scope of this thesis, which was focused on assessing whether anxiety could be contributing to negative attribution bias in a TBI population. There was no evidence that anxiety was associated with negative attribution bias on the Bristol ERT. Future studies should consider independent associations with aggression and also irritability, which is commonly reported after mild TBI (Deveney et al., 2019; Prince & Bruhns, 2017). Jorna and colleagues (2021) recently reported increased anger misattribution on FEEST in an acquired brain injury sample compared to controls. Even though the effect size was small, they found evidence of an association with levels of socially appropriate behaviour reported by other people. This highlights the importance

of further research in this area and indicates that anger misattributions on the Bristol ERT could be a useful indicator of negative attribution bias. Furthermore, by adding feedback to the Bristol EBT it can be used as a bias modification task to reduce levels of aggression, which could be explored as an intervention in a TBI population (Penton-Voak et al., 2017; Penton-Voak et al., 2013).

The fact that there is evidence for an association between performance on the Bristol ERT and moderate to severe TBI, but not mild TBI confirms that it is important to consider severity of injury when assessing emotion recognition in a TBI population. Murphy and colleagues (2021) have suggested that there is no difference in the degree of emotion recognition impairment between moderate and severe TBI, but that this may be different for mild TBI. In this thesis mild TBI was defined as having a head injury with loss of consciousness. Whilst it is a commonly used indicator of mild TBI (Ruff et al., 2009), it does not provide much of an indication about the extent of possible brain damage. Gasquoine (2020) has argued that Post Traumatic Amnesia (PTA) is a better indicator that mild TBI has occurred potentially because it suggests greater disruption of brain function than loss of consciousness. Notably, in Chapter 5, a trend towards decreased emotion recognition was identified for participants with mild TBI who reported having PTA. However, there was no evidence of this association once age and ability to perceive faces was controlled for. Treating mild TBI as a homogeneous group may be obscuring decreased emotion recognition in a sub-group of people with mild TBI. It would be better to include a more continuous measure of mild TBI when assessing emotion recognition in this population. Alternatively, associations with emotion recognition should be assessed in sub-groups of mild TBI, such as those identified by Si and colleagues (2018).

7.4 Aim Three - Anxiety and Emotion Recognition after Traumatic Brain Injury

The third aim of this thesis was to investigate whether anxiety could be moderating changes in emotion recognition in a TBI population. Having a TBI is associated with increased levels of anxiety and higher likelihood of having an anxiety disorder (Osborn et al., 2016; Ponsford et al., 2018). To evaluate whether anxiety could be contributing to or even explain changes in emotion recognition after TBI a continuous measure of clinical anxiety (DASS anxiety subscale) was included in the cross-sectional studies presented in Chapter 4 and Chapter 5. In Chapter 6 associations between mild TBI, emotion recognition and Generalised Anxiety Disorder (GAD) as well as reports of

recent anxiety were explored in ALSPAC. In line with the current literature, there was evidence of increased levels of anxiety on the DASS in the TBI groups compared to controls and having a mild TBI was associated with an increased likelihood of having GAD in the ALSPAC sample. Despite the evidence for increased anxiety in the TBI samples, the DASS anxiety scores and a diagnosis of GAD were not associated with emotion recognition performance on the Bristol ERT. Based on these findings it seems unlikely that anxiety is moderating changes in emotion recognition observed in a TBI population.

Although having a diagnosis of GAD was not associated with performance on the Bristol ERT, reports of recent anxiety in ALSPAC (having anxiety in the past week) was associated with better performance on the Bristol ERT. Attwood and colleagues (2017) similarly found that trait anxiety measured using the State-Trait Anxiety Inventory (Spielberger, 1983) was associated with increased scores on the Bristol ERT, although other studies have not found evidence for this association (Cooper et al., 2008; Dyer et al., 2022). Knowles and Olatunji (2020) have argued that the trait scale on the State-Trait Anxiety Inventory is a measure of negative affect, not specifically anxiety. In fact trait scale on the State-Trait Anxiety Inventory is most highly correlated with the depression subscale on the DASS (Antony et al., 1998). Interestingly, the depression and stress subscales on the DASS were also associated with increased emotion recognition in this thesis (Chapter 5). These results suggest that general negative affect might be associated with an increased performance on the Bristol ERT, and it is plausible that the question about recent anxiety in ALSPAC (Chapter 6) was actually a measure of general negative affect.

One of the reasons that clinical levels of anxiety was investigated as a potential moderator of performance on the Bristol ERT after TBI was that Attwood and colleagues (2017) reported decreased performance on the task when anxiety was induced. They also reported negative bias on the Bristol EBT (Happy - Angry) in participants with state anxiety. These results were recently replicated by Dyer and colleagues (2022), although negative attribution bias was only observed when state anxiety was induced in participants with high trait anxiety. There was no evidence clinical anxiety being associated with performance on the Bristol ERT or EBT tasks in the studies presented in Chapter 4 and Chapter 5. The DASS is not considered a measure of trait anxiety (Elwood et al., 2012), but as it asks about symptoms of anxiety experienced in the past seven days it also does not capture whether participants are in an acute state of anxiety. It

stands to reason, that scores on the DASS anxiety subscale indicate whether people are more likely to regularly experience periods of acute anxiety. Manierka and colleagues (2021) have found that mood fluctuations are associated with substantial changes in performance on emotion recognition tasks in a healthy population. This suggests that if participants were not in an acute state of anxiety at the time of testing scores on the Bristol ERT would not be impaired but that they could still regularly experience difficulties in emotion recognition due to anxiety. Further research is needed to understand the potential impact of state anxiety on emotion recognition in a TBI population.

Another consideration is that measures of anxiety in this thesis were all based on self-report. Suslow and colleagues (2019) have reported dissociations between implicit and explicit measures of anxiety in a healthy population. Furthermore, having a TBI is often associated with a lack of insight in relation of self-awareness and reporting of symptoms (Azouvi et al., 2017). Studies have shown that performance on emotion recognition tasks is associated with other report, but not self-report, of difficulties in a TBI population (Jorna et al., 2021; Spikman, Milders, et al., 2013). It is possible that the lack of association between performance on the Bristol ERT and anxiety was due to difficulties in awareness and reporting of anxiety. Although the fact that there was evidence of increased anxiety in the TBI groups compared to controls would suggest that the TBI participants were able to identify feelings of anxiety. Nevertheless, it is worth exploring whether clinician or family reported anxiety are associated with performance on the Bristol ERT in a TBI population.

7.5 Novel Contributions

This thesis provides the first comprehensive evaluation of psychometric properties for the Bristol ERT. Understanding these properties is an important step towards developing the Bristol ERT for use in a clinical setting (Kelly et al., 2017; Kessels, 2019). Insights gained have led to concrete suggestions about how to improve measurement of emotion recognition using the Bristol ERT. These considerations are also important for inferences about emotion recognition made in a research setting, especially as this type of research is difficult to fund and publish (Vitoratou & Pickles, 2017). For example, reliability estimates reported in this thesis could be used to inform sample size calculations to help avoid low statistical power when measuring emotion recognition using the Bristol ERT (Cooper et al., 2017; Parsons et al., 2019). It is also the

first time that the Bristol ERT has been used to assess emotion recognition in a TBI population. Results indicate that performance on the Bristol ERT after moderate to severe TBI is comparable to performance on other emotion recognition tasks using facial images in a TBI population. Furthermore, few studies to date have investigated emotion recognition after mild TBI (Calvillo & Irimia, 2020; Murphy et al., 2021). Two studies assessing emotion recognition after mild TBI were conducted as part of this thesis to address this gap in the literature. The lack of association observed between performance on the Bristol ERT and mild TBI suggests that emotion recognition deficits reported in moderate to severe TBI samples are not present in a mild TBI population. Using different methodologies to investigate associations between emotion recognition and mild TBI increases confidence in the observations made, as triangulation helps reduce bias associated with a single research approach. Finally, this thesis also contributes to our understanding of how anxiety may be linked to changes in emotion recognition associated with having a TBI.

7.6 Future Directions

7.6.1 Developing Bristol ERT for use in a healthcare setting

Kessels (2019) has outlined importance of and challenges of adapting computerised experimental tasks for use in neuropsychological assessment. One of the challenges is that tasks lack information about psychometric properties and many tasks used in neurological samples have not been sufficiently validated in the population of interest (D'Souza et al., 2019; Howieson, 2019). The work presented in this thesis addresses this issue and provides a good basis for further development. Several recommendations regarding the development of the Bristol ERT have already been discussed, once implemented further studies will be needed to assess psychometric properties in both healthy and neurological populations. It will also be important to establish age specific normative data for the Bristol ERT in both male and female samples. Ideally this should be done once the task is deemed suitable for the intended use so that no changes are required once normative data is established (Kessels, 2019). Normative data is useful because it will allow for percentiles to be calculated to allow for interpretation of individual emotion recognition scores.

Digitalisation for assessment and treatment in a healthcare setting is a big area of development, the importance of which has increased further since the COVID-19 (Kwasnicka et al., 2022). A benefit of developing the Bristol ERT as a tool for

neuropsychological assessment is that it is already digitally available as part of CANTAB. An interesting next step could be to adapt the way the Bristol ERT is presented by using ecological momentary assessments instead of relying on a single testing session. Depp and colleagues (2021) have successfully used mobile phones to present an emotion recognition task repeatedly over the course of 10 days. They reported that mean performance across repeated presentations was highly correlated with a lab-based version of the same task, suggesting it could be a viable form of assessment. Considering the evidence for changes in performance on the Bristol ERT related to state anxiety (Attwood et al., 2017; Dyer et al., 2022), this form of multiple assessment together with momentary reports of anxiety could provide valuable insights.

7.6.2 Establishing clinical significance

The results in this thesis clearly show that people with moderate to severe TBI perform worse than neurologically healthy controls on the Bristol ERT. The size of this effect is considered large according to Cohen's widely used thresholds, but these thresholds are arbitrary and do not tell us whether the deficit corresponds with impaired functioning in everyday life (Lakens, 2013b). Measuring emotion recognition in clinical practice is only valuable if it can be used to make functional predication (Kessels, 2019), for example the maintenance or quality of social relationships. An important next step will be to establish a clinically significant difference for emotion recognition deficits on the Bristol ERT, in other words the smallest effect size of interest (Anvari & Lakens, 2021). The smallest effect size of interest might vary for different functions of interest, so it will be important to define what the functional outcomes of interest are and evaluate effect sizes for each of these. Anvari and Lakens (2021) have published a guide on how to use an anchor-based method to establish an effect size of interest, which could be used as a guide for future studies. Given the evidence that performance on emotion recognition tasks is associated with other (not self) report of functional difficulties (Jorna et al., 2021; Spikman, Milders, et al., 2013), it would be interesting establish effect sizes of interest for both. For example, a future study could investigate whether people with TBI identified by clinicians as presenting with social difficulties differ in performance on the Bristol ERT compared to people with similar injury severity without reported difficulties. Establishing the smallest effect size of interest on the Bristol ERT could also be helpful to assess whether having mild TBI is associated with meaningful change in emotion recognition, or rather investigate the absence of an association (Lakens et al., 2020). Having clinically meaningful reference point will further

be beneficial for the development and comparison of emotion perception treatments (Cassel et al., 2019).

7.6.3 Emotion recognition and Traumatic Brain Injury

There is still a lot that we do not understand about emotion recognition in a TBI population. The fact that participants with moderate to severe TBI show decreased emotion recognition compared to neurologically healthy controls, but participants with mild TBI do not highlights the importance of thinking about severity of injury when assessing emotion recognition. Instead of using the diagnostic categories of mild, moderate, or severe TBI when assessing emotion recognition, it would be useful to investigate whether individual symptoms used to inform diagnosis are predictors of performance on emotion recognition tasks. For example, investigating whether length of PTA or reporting of persisting symptoms (e.g. headaches or dizziness) in a mild TBI population are associated with performance on the Bristol ERT. Data driven categories of mild TBI (Si et al., 2018) could also provide useful insights about the association between emotion recognition and TBI. Finally, as brain imaging techniques like Diffuse Tensor Imaging become more widely available the degree of axonal damage after TBI could also be a useful predicator of emotion recognition difficulties (Yassin et al., 2017).

Further research is also needed to investigate whether the decrease in emotion recognition observed after moderate to severe TBI are causally related (Theadom et al., 2019). Including repeated measures of emotion recognition and TBI as part of prospective cohort studies would allow for investigation of causal links. Prospective cohort studies are incredibly time and cost intensive (Mann, 2003) but as methods for linking healthcare data to existing cohort studies improves (Harron et al., 2020), these types of investigation will hopefully become more feasible. A better understanding of causal pathways could also benefit the development of interventions for people struggling with emotion recognition difficulties.

References

Abbruzzese, L., Magnani, N., Robertson, I. H., & Mancuso, M. (2019). Age and Gender Differences in Emotion Recognition. *Frontiers in Psychology*, *10*, Article 2371. https://doi.org/10.3389/fpsyg.2019.02371

Abushalbaq, O. M., Khdour, H. Y., Hamza, E. G. A., Moustafa, A. A., & Herzallah, M. M. (2021). Investigating Principal Working Memory Features in Generalized, Panic, and Social Anxiety Spectrum Disorders. *Frontiers in Psychiatry*, *12*, Article 701412. https://doi.org/10.3389/fpsyt.2021.701412

Adachi, P., & Willoughby, T. (2015). Interpreting effect sizes when controlling for stability effects in longitudinal autoregressive models: Implications for psychological science [Article]. *European Journal of Developmental Psychology*, *12*(1), 116-128. https://doi.org/10.1080/17405629.2014.963549

Adams, T., Pounder, Z., Preston, S., Hanson, A., Gallagher, P., Harmer, C. J., & McAllister-Williams, R. H. (2016). Test-retest reliability and task order effects of emotional cognitive tests in healthy subjects. *Cogn Emot*, *30*(7), 1247-1259. https://doi.org/10.1080/02699931.2015.1055713

Adolphs, R. (2002a). Neural systems for recognizing emotion. *Current Opinion in Neurobiology*, *12*(2), 169-177. https://doi.org/10.1016/s0959-4388(02)00301-x

Adolphs, R. (2002b). Recognizing emotion from facial expressions: psychological and neurological mechanisms. *Behavioral cognitive neuroscience reviews*, *1*(1), 21-62.

Adolphs, R. (2010). What does the amygdala contribute to social cognition? *Year in Cognitive Neuroscience 2010*, *1191*, 42-61. https://doi.org/10.1111/j.1749-6632.2010.05445.x

Allain, P., Togher, L., & Azouvi, P. (2019). Social cognition and traumatic brain injury: current knowledge [Editorial Material]. *Brain Injury*, *33*(1), 1-3. https://doi.org/10.1080/02699052.2018.1533143

Alway, Y., Gould, K. R., Johnston, L., McKenzie, D., & Ponsford, J. (2016). A prospective examination of Axis I psychiatric disorders in the first 5 years following moderate to severe traumatic brain injury. *Psychological Medicine*, *46*(6), 1331-1341. https://doi.org/10.1017/s0033291715002986

Antony, M. M., Bieling, P. J., Cox, B. J., Enns, M. W., & Swinson, R. P. (1998). Psychometric properties of the 42-item and 21-item versions of the Depression Anxiety Stress Scales in clinical groups and a community sample [Article]. *Psychological Assessment*, *10*(2), 176-181. https://doi.org/10.1037/1040-3590.10.2.176

Anvari, F., & Lakens, D. (2018). The replicability crisis and public trust in psychological science. *Comprehensive Results in Social Psychology*, *3*(3), 266-286.

Anvari, F., & Lakens, D. (2021). Using anchor-based methods to determine the smallest effect size of interest. *Journal of Experimental Social Psychology*, *96*, Article 104159. https://doi.org/10.1016/j.jesp.2021.104159

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2019). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods*, *52*(1), 388-407.

Ashman, T. A., Spielman, L. A., Hibbard, M. R., Silver, J. M., Chandna, T., & Gordon, W. A. (2004). Psychiatric challenges in the first 6 years after traumatic brain injury: Cross-sequential analyses of axis I disorders. *Archives of Physical Medicine and Rehabilitation*, *85*(4), S36-S42. https://doi.org/10.1016/j.apmr.2003.08.117

Association, A. E. R., Association, A. P., & Education, N. C. o. M. i. (1999). *Standards for educational and psychological testing*. American Educational Research Association.

Association, A. E. R., Association, A. P., & Education, N. C. o. M. i. (2014). *Standards for educational and psychological testing*. American Educational Research Association.

Atkinson, A. P., Dittrich, W. H., Gemmell, A. J., & Young, A. W. (2004). Emotion perception from dynamic and static body expressions in point-light and full-light displays [Article]. *Perception*, *33*(6), 717-746. https://doi.org/10.1068/p5096

Attwood, A. S., Easey, K. E., Dalili, M. N., Skinner, A. L., Woods, A., Crick, L., Ilett, E., Penton-Voak, I. S., & Munafo, M. R. (2017). State anxiety and emotional face recognition in healthy volunteers [Article]. *Royal Society Open Science*, *4*(5), 16, Article Unsp 160855. https://doi.org/10.1098/rsos.160855

Avery, S. N., VanDerKlok, R. M., Heckers, S., & Blackford, J. U. (2016). Impaired face recognition is associated with social inhibition. *Psychiatry research*, *236*, 53-57.

Aviezer, H., Hassin, R. R., Ryan, J., Grady, C., Susskind, J., Anderson, A., Moscovitch, M., & Bentin, S. (2008). Angry, disgusted, or afraid? Studies on the malleability of emotion perception [Article]. *Psychological Science*, *19*(7), 724-732. https://doi.org/10.1111/j.1467-9280.2008.02148.x

Azouvi, P., Arnould, A., Dromer, E., & Vallat-Azouvi, C. (2017). Neuropsychology of traumatic brain injury: An expert overview [Article]. *Revue Neurologique*, *173*(7-8), 461-472. https://doi.org/10.1016/j.neurol.2017.07.006

Babbage, D. R., Yim, J., Zupan, B., Neumann, D., Tomita, M. R., & Willer, B. (2011). Meta-analysis of facial affect recognition difficulties after traumatic brain injury. *Neuropsychology*, *25*(3), 277-285. https://doi.org/10.1037/a0021908

Bagby, R. M., Parker, J. D. A., & Taylor, G. J. (1994). THE 20-ITEM TORONTO-ALEXITHYMIA-SCALE .1. ITEM SELECTION AND CROSS-VALIDATION OF THE FACTOR STRUCTURE [Article]. *Journal of psychosomatic research*, *38*(1), 23-32. https://doi.org/10.1016/0022-3999(94)90005-1

Bagby, R. M., & Taylor, G. J. (1997). Affect dysregulation and alexithymia. In G. J. Taylor, J. D. A. Parker, & R. M. Bagby (Eds.), *Disorders of Affect Regulation: Alexithymia in Medical and Psychiatric Illness* (pp. 26-45). Cambridge University Press. https://doi.org/DOI: 10.1017/CBO9780511526831.005

Bagby, R. M., Taylor, G. J., & Parker, J. D. A. (1994). THE 20-ITEM TORONTO-ALEXITHYMIA-SCALE .2. CONVERGENT, DISCRIMINANT, AND CONCURRENT VALIDITY [Article]. *Journal of psychosomatic research*, *38*(1), 33-40. https://doi.org/10.1016/0022-3999(94)90006-x

Bamford, S., Penton-Voak, I., Pinkney, V., Baldwin, D. S., Munafo, M. R., & Garner, M. (2015). Early effects of duloxetine on emotion recognition in healthy volunteers. *J Psychopharmacol*, *29*(5), 634-641. https://doi.org/10.1177/0269881115570085

Band, G. P. H., van der Molen, M. W., & Logan, G. D. (2003). Horse-race model simulations of the stop-signal procedure. *Acta Psychologica*, *112*(2), 105-142, Article Pii s001-6918(02)00079-3. https://doi.org/10.1016/s0001-6918(02)00079-3

Bannigan, K., & Watson, R. (2009). Reliability and validity in a nutshell. *Journal of clinical nursing*, *18*(23), 3237-3243. https://onlinelibrary.wiley.com/doi/full/10.1111/j.1365-2702.2009.02939.x

Barker-Collo, S., Theadom, A., Jones, K., Starkey, N., Kahan, M., & Feigin, V. (2018). Depression and anxiety across the first 4 years after mild traumatic brain injury: findings from a community-based study. *Brain Injury*, *32*(13-14), 1651-1658. https://doi.org/10.1080/02699052.2018.1540797

Baugh, F. (2002). Correcting effect sizes for score reliability: A reminder that measurement and substantive issues are linked inextricably. *Educational and*

psychological measurement, *62*(2), 254-263. https://doi.org/10.1177/0013164402062002004
Beer, J. S., & Ochsner, K. N. (2006). Social cognition: A multi level analysis. *Brain Research*, *1079*, 98-105. https://doi.org/10.1016/j.brainres.2006.01.002
Bell, R. J. (2020). Why do we need cohort studies? *Climacteric*, *23*(4), 321-322. https://doi.org/10.1080/13697137.2020.1764526
Berchtold, A. (2016). Test–retest: agreement or reliability? *Methodological Innovations*, *9*, 2059799116672875.
Bigler, E. D. (2007). Anterior and middle cranial fossa in traumatic brain injury: Relevant neuroanatomy and neuropathology in the study of neuropsychological outcome. *Neuropsychology*, *21*(5), 515-531. https://doi.org/10.1037/0894-4105.21.5.515
Bigler, E. D. (2013). Neuroimaging biomarkers in mild traumatic brain injury (mTBI). *Neuropsychol Rev*, *23*(3), 169-209. https://doi.org/10.1007/s11065-013-9237-2
Biszak, A. M., & Babbage, D. R. (2014). Facial affect recognition difficulties in traumatic brain injury rehabilitation services. *Brain Injury*, *28*(1), 97-104. https://doi.org/10.3109/02699052.2013.856475
Bland, A. R., Roiser, J. P., Mehta, M. A., Schei, T., Boland, H., Campbell-Meiklejohn, D. K., Emsley, R. A., Munafo, M. R., Penton-Voak, I. S., Seara-Cardoso, A., Viding, E., Voon, V., Sahakian, B. J., Robbins, T. W., & Elliott, R. (2016). EMOTICOM: A Neuropsychological Test Battery to Evaluate Emotion, Motivation, Impulsivity, and Social Cognition. *Front Behav Neurosci*, *10*, 25. https://doi.org/10.3389/fnbeh.2016.00025
Bohorquez-Montoya, L., Espana, L. Y., Nader, A. M., Furger, R. E., Mayer, A. R., & Meier, T. B. (2020). Amygdala response to emotional faces in adolescents with persistent post-concussion symptoms. *Neuroimage-Clinical*, *26*, Article 102217. https://doi.org/10.1016/j.nicl.2020.102217
Bonora, A., Benuzzi, F., Monti, G., Mirandola, L., Pugnaghi, M., Nichelli, P., & Meletti, S. (2011). Recognition of emotions from faces and voices in medial temporal lobe epilepsy. *Epilepsy & Behavior*, *20*(4), 648-654. https://doi.org/10.1016/j.yebeh.2011.01.027
Bowden, S. C. (2017). *Neuropsychological assessment in the age of evidence-based practice: diagnostic and treatment evaluations*. Oxford University Press.
Bowden, S. C., Petrauskas, V. M., Bardenhagen, F. J., Meade, C. E., & Simpson, L. C. (2013). Exploring the Dimensionality of Digit Span [Article]. *Assessment*, *20*(2), 188-198. https://doi.org/10.1177/1073191112457016
Boyd, A., Golding, J., Macleod, J., Lawlor, D. A., Fraser, A., Henderson, J., Molloy, L., Ness, A., Ring, S., & Davey Smith, G. (2013). Cohort profile: the 'children of the 90s'—the index offspring of the Avon Longitudinal Study of Parents and Children. *International journal of epidemiology*, *42*(1), 111-127.
Bredemeier, K., & Berenbaum, H. (2013). Cross-Sectional and Longitudinal Relations between Working Memory Performance and Worry. *Journal of Experimental Psychopathology*, *4*(4), 420-434. https://doi.org/10.5127/jep.032212
Brown, A. W., Pretz, C. R., Bell, K. R., Hammond, F. M., Arciniegas, D. B., Bodien, Y. G., Dams-O'Connor, K., Giacino, J. T., Hart, T., Johnson-Greene, D., Kowalski, R. G., Walker, W. C., Weintraub, A., & Zafonte, R. (2019). Predictive utility of an adapted Marshall head CT classification scheme after traumatic brain injury. *Brain Injury*, *33*(5), 610-617. https://doi.org/10.1080/02699052.2019.1566970
Bruton, A., Conway, J. H., & Holgate, S. T. (2000). Reliability: what is it, and how is it measured? *Physiotherapy*, *86*(2), 94-99.

Burton, A. M., White, D., & McNeill, A. (2010). The Glasgow Face Matching Test [Article]. *Behavior Research Methods*, *42*(1), 286-291. https://doi.org/10.3758/brm.42.1.286

Buss, A. H., & Perry, M. (1992). THE AGGRESSION QUESTIONNAIRE [Article]. *Journal of Personality and Social Psychology*, *63*(3), 452-459. https://doi.org/10.1037/0022-3514.63.3.452

Button, K. S., Ioannidis, J. P., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S., & Munafò, M. R. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, *14*(5), 365-376.

Byom, L., Duff, M., Mutlu, B., & Turkstra, L. (2019). Facial emotion recognition of older adults with traumatic brain injury. *Brain Injury*, *33*(3), 322-332. https://doi.org/10.1080/02699052.2018.1553066

Calamia, M., Markon, K., & Tranel, D. (2012). Scoring higher the second time around: meta-analyses of practice effects in neuropsychological assessment. *The Clinical Neuropsychologist*, *26*(4), 543-570.

Callahan, B. L., Ueda, K., Sakata, D., Plamondon, A., & Murai, T. (2011). Liberal bias mediates emotion recognition deficits in frontal traumatic brain injury. *Brain and Cognition*, *77*(3), 412-418. https://doi.org/10.1016/j.bandc.2011.08.017

Calvillo, M., & Irimia, A. (2020). Neuroimaging and Psychometric Assessment of Mild Cognitive Impairment After Traumatic Brain Injury. *Frontiers in Psychology*, *11*, Article 1423. https://doi.org/10.3389/fpsyg.2020.01423

Calvo, M. G., Avero, P., Fernández-Martín, A., & Recio, G. (2016). Recognition thresholds for static and dynamic emotional faces. *Emotion*, *16*(8), 1186.

Camargo, S. L., Herrera, A. N., & Traynor, A. (2018). Looking for a Consensus in the Discussion About the Concept of Validity A Delphi Study [Article]. *Methodology-European Journal of Research Methods for the Behavioral and Social Sciences*, *14*(4), 146-155. https://doi.org/10.1027/1614-2241/a000157

Campbell, D. T., & Fiske, D. W. (1959). CONVERGENT AND DISCRIMINANT VALIDATION BY THE MULTITRAIT-MULTIMETHOD MATRIX. *Psychological Bulletin*, *56*(2), 81-105. https://doi.org/10.1037/h0046016

Carroll, L. J., Cassidy, J. D., Holm, L., Kraus, J., & Coronado, V. G. (2004). Methodological issues and research recommendations for mild traumatic brain injury: The WHO Collaborating Centre Task Force on Mild Traumatic Brain Injury [Review]. *Journal of Rehabilitation Medicine*, *36*, 113-125. https://doi.org/10.1080/16501960410023877

Cassel, A., McDonald, S., Kelly, M., & Togher, L. (2019). Learning from the minds of others: A review of social cognition treatments and their relevance to traumatic brain injury [Review]. *Neuropsychological Rehabilitation*, *29*(1), 22-55. https://doi.org/10.1080/09602011.2016.1257435

Castro, V. L., Cheng, Y. H., Halberstadt, A. G., & Gruhn, D. (2016). EUReKA! A Conceptual Model of Emotion Understanding. *Emotion Review*, *8*(3), 258-268. https://doi.org/10.1177/1754073915580601

Cecilione, J. L., Rappaport, L. M., Verhulst, B., Carney, D. M., Blair, R. J. R., Brotman, M. A., Leibenluft, E., Pine, D. S., Roberson-Nay, R., & Hettema, J. M. (2017). Test-Retest Reliability of the Facial Expression Labeling Task. *Psychological Assessment*, *29*(12), 1537-1542. https://doi.org/10.1037/pas0000439

Celeghin, A., Diano, M., Bagnis, A., Viola, M., & Tamietto, M. (2017). Basic Emotions in Human Neuroscience: Neuroimaging and Beyond. *Frontiers in Psychology*, *8*, Article 1432. https://doi.org/10.3389/fpsyg.2017.01432

Champely, S. (2020). *pwr package: Basic Functions for Power Analysis*. In (Version 1.3-0) https://github.com/heliosdrm/pwr

Cicchetti, D. V. (2001). The precision of reliability and validity estimates re-visited: Distinguishing between clinical and statistical significance of sample size requirements. *Journal of Clinical and Experimental Neuropsychology*, *23*(5), 695-700. https://doi.org/10.1076/jcen.23.5.695.1249

Cnossen, M. C., Winkler, E. A., Yue, J. K., Okonkwo, D. O., Valadka, A. B., Steyerberg, E. W., Lingsma, H. F., Manley, G. T., & Investigators, T.-T. (2017). Development of a Prediction Model for Post-Concussive Symptoms following Mild Traumatic Brain Injury: A TRACK-TBI Pilot Study. *Journal of Neurotrauma*, *34*(16), 2396-2409. https://doi.org/10.1089/neu.2016.4819

Colling, L. J., & Szucs, D. (2021). Statistical Inference and the Replication Crisis. *Review of Philosophy and Psychology*, *12*(1), 121-147. https://doi.org/10.1007/s13164-018-0421-4

Colonnelle, V., Russo, P. M., & Mattarozzi, K. (2019). First Impression Misleads Emotion Recognition. *Frontiers in Psychology*, *10*, Article 527. https://doi.org/10.3389/fpsyg.2019.00527

Cook, D. A., & Beckman, T. J. (2006). Current concepts in validity and reliability for psychometric instruments: theory and application. *The American journal of medicine*, *119*(2), 166. e167-166. e116.

Cooper, R. M., Rowe, A. C., & Penton-Voak, I. S. (2008). The role of trait anxiety in the recognition of emotional facial expressions. *J Anxiety Disord*, *22*(7), 1120-1127. https://doi.org/10.1016/j.janxdis.2007.11.010

Cooper, S. R., Gonthier, C., Barch, D. M., & Braver, T. S. (2017). The role of psychometrics in individual differences research in cognition: A case study of the AX-CPT. *Frontiers in Psychology*, *8*, 1482.

Crawford, J. R., & Henry, J. D. (2003). The Depression Anxiety Stress Scales (DASS): Normative data and latent structure in a large non-clinical sample [Article]. *British Journal of Clinical Psychology*, *42*, 111-131. https://doi.org/10.1348/014466503321903544

Crivelli, C., & Fridlund, A. J. (2019). Inside-Out: From Basic Emotions Theory to the Behavioral Ecology View. *Journal of Nonverbal Behavior*, *43*(2), 161-194. https://doi.org/10.1007/s10919-019-00294-2

Croker, V., & McDonald, S. (2005). Recognition of emotion from facial expression following traumatic brain injury [Article]. *Brain Injury*, *19*(10), 787-799. https://doi.org/10.1080/02699050500110033

Cronbach, L. J., & Meehl, P. E. (1955). CONSTRUCT VALIDITY IN PSYCHOLOGICAL TESTS [Article]. *Psychological Bulletin*, *52*(4), 281-302. https://doi.org/10.1037/h0040957

Curran, C. A., Ponsford, J., & Crowe, S. (2000). Coping strategies and emotional outcome following traumatic brain injury: A comparison with orthopedic patients [Article]. *Journal of Head Trauma Rehabilitation*, *15*(6), 1256-1274. https://doi.org/10.1097/00001199-200012000-00006

D'Souza, A., Mollayeva, S., Pacheco, N., Javed, F., Colantonio, A., & Mollayeva, T. (2019). Measuring Change Over Time: A Systematic Review of Evaluative Measures of Cognitive Functioning in Traumatic Brain Injury [Review]. *Frontiers in Neurology*, *10*, 24, Article 353. https://doi.org/10.3389/fneur.2019.00353

Dahm, J., & Ponsford, J. (2015). Predictors of global functioning and employment 10 years following traumatic brain injury compared with orthopaedic injury. *Brain Injury*, *29*(13-14), 1539-1546. https://doi.org/10.3109/02699052.2015.1075141

Dalili, M. N., Penton-Voak, I. S., Harmer, C. J., & Munafo, M. R. (2015). Meta-analysis of emotion recognition deficits in major depressive disorder. *Psychol Med*, *45*(6), 1135-1144. https://doi.org/10.1017/S0033291714002591

Darke, H., Cropper, S. J., & Carter, O. (2019). A Novel Dynamic Morphed Stimuli Set to Assess Sensitivity to Identity and Emotion Attributes in Faces [Article]. *Frontiers in Psychology*, *10*, 18, Article 757. https://doi.org/10.3389/fpsyg.2019.00757

De Schryver, M., Hughes, S., Rosseel, Y., & De Houwer, J. (2016). Unreliable yet still replicable: A comment on LeBel and Paunonen (2011). *Frontiers in Psychology*, *6*, 2039.

de Sousa, A., McDonald, S., Rushby, J., Li, S., Dimoska, A., & James, C. (2010). Why don't you feel how I feel? Insight into the absence of empathy after severe Traumatic Brain Injury. *Neuropsychologia*, *48*(12), 3585-3595. https://doi.org/10.1016/j.neuropsychologia.2010.08.008

Delmonico, R. L., Theodore, B. R., Sandel, M. E., Armstrong, M. A., & Camicia, M. (2021). Prevalence of depression and anxiety disorders following mild traumatic brain injury [https://doi.org/10.1002/pmrj.12657]. *PM&R*, *n/a*(n/a). https://doi.org/https://doi.org/10.1002/pmrj.12657

Demenescu, L. R., Kortekaas, R., den Boer, J. A., & Aleman, A. (2010). Impaired Attribution of Emotion to Facial Expressions in Anxiety and Major Depression [Article]. *PLoS One*, *5*(12), 5, Article e15058. https://doi.org/10.1371/journal.pone.0015058

Depp, C. A., Kamarsu, S., Filip, T. F., Parrish, E. M., Harvey, P. D., Granholm, E. L., Chalker, S., Moore, R. C., & Pinkham, A. (2021). Ecological momentary facial emotion recognition in psychotic disorders. *Psychological Medicine*, 1-9.

DeVellis, R. F. (2006). Classical test theory. *Medical Care*, *44*(11), S50-S59. https://doi.org/10.1097/01.mlr.0000245426.10853.30

Deveney, C. M., Chen, S. H., Wilmer, J. B., Zhao, V., Schmidt, H. B., & Germine, L. (2018). How generalizable is the inverse relationship between social class and emotion perception? *PLoS One*, *13*(10), Article e0205949. https://doi.org/10.1371/journal.pone.0205949

Deveney, C. M., Stoddard, J., Evans, R. L., Chavez, G., Harney, M., & Wulff, R. A. (2019). On defining irritability and its relationship to affective traits and social interpretations. *Personality and Individual Differences*, *144*, 61-67. https://doi.org/10.1016/j.paid.2019.02.031

Dewan, M. C., Rattani, A., Gupta, S., Baticulon, R. E., Hung, Y. C., Punchak, M., Agrawal, A., Adeleye, A. O., Shrime, M. G., Rubiano, A. M., Rosenfeld, J. V., & Park, K. B. (2019). Estimating the global incidence of traumatic brain injury. *Journal of Neurosurgery*, *130*(4), 1080-1097. https://doi.org/10.3171/2017.10.jns17352

Dikmen, S., Machamer, J., Fann, J. R., & Temkin, N. R. (2010). Rates of symptom reporting following traumatic brain injury. *Journal of the International Neuropsychological Society*, *16*(3), 401-411. https://doi.org/10.1017/s1355617710000196

Dodge, K. A. (2006). Translational science in action: Hostile attributional style and the development of aggressive behavior problems. *Development and psychopathology*, *18*(3), 791-814. https://doi.org/10.1017/s0954579406060391

Doty, T. J., Japee, S., Ingvar, M., & Ungerleider, L. G. (2013). Fearful Face Detection Sensitivity in Healthy Adults Correlates With Anxiety-Related Traits. *Emotion*, *13*(2), 183-188. https://doi.org/10.1037/a0031373

Draper, K., & Ponsford, J. (2008). Cognitive functioning ten years following traumatic brain injury and rehabilitation. *Neuropsychology*, *22*(5), 618-625. https://doi.org/10.1037/0894-4105.22.5.618

Duchaine, B. C., Parker, H., & Nakayama, K. (2003). Normal recognition of emotion in a prosopagnosic. *Perception*, *32*(7), 827-838. https://doi.org/10.1068/p5067

Dukes, D., Clement, F., Audrin, C., & Mortillaro, M. (2017). Looking beyond the static face in emotion recognition: The informative case of interest. *Visual Cognition*, *25*(4-6), 575-588. https://doi.org/10.1080/13506285.2017.1341441

Dunning, D. L., Westgate, B., & Adlam, A. L. R. (2016). A Meta-Analysis of Working Memory Impairments in Survivors of Moderate-to-Severe Traumatic Brain Injury. *Neuropsychology*, *30*(7), 811-819. https://doi.org/10.1037/neu0000285

Dyer, M. L., Attwood, A. S., Penton-Voak, I. S., & Munafo, M. R. (2022). The role of state and trait anxiety in the processing of facial expressions of emotion. *Royal Society Open Science*, *9*(1), Article 210056. https://doi.org/10.1098/rsos.210056

Eastwood, A. P. R., Penton-Voak, I. S., Munafò, M. R., & Attwood, A. S. (2020). Effects of acute alcohol consumption on emotion recognition in high and low trait aggressive drinkers. *Journal of Psychopharmacology*, *34*(11), 1226-1236. https://doi.org/10.1177/0269881120922951

Ekman, P. (1992). AN ARGUMENT FOR BASIC EMOTIONS [Article]. *Cognition & Emotion*, *6*(3-4), 169-200. https://doi.org/10.1080/02699939208411068

Ekman, P., & Cordaro, D. (2011). What is meant by calling emotions basic. *Emotion Review*, *3*(4), 364-370.

Ekman, P., & Friesen, W. V. (1978). *Facial action coding system*. Consulting Psychologists Press.

Ekman, P., Friesen, W. V., Osullivan, M., Chan, A., Diacoyannitarlatzis, I., Heider, K., Krause, R., Lecompte, W. A., Pitcairn, T., Riccibitti, P. E., Scherer, K., Tomita, M., & Tzavaras, A. (1987). UNIVERSALS AND CULTURAL-DIFFERENCES IN THE JUDGMENTS OF FACIAL EXPRESSIONS OF EMOTION [Article]. *Journal of Personality and Social Psychology*, *53*(4), 712-717. https://doi.org/10.1037/0022-3514.53.4.712

Elwood, L. S., Wolitzky-Taylor, K., & Olatunji, B. O. (2012). Measurement of anxious traits: a contemporary review and synthesis. *Anxiety, Stress & Coping*, *25*(6), 647-666.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175-191.

Ferreira, C. D., & Torro-Alves, N. (2016). Facial emotion recognition in aging: a systematic review. *Universitas Psychologica*, *15*(5). https://doi.org/10.11144/Javeriana.upsy15-5.refe

Fisher, R. A. (1925). Theory of statistical estimation. Mathematical Proceedings of the Cambridge Philosophical Society,

Fordington, S., & Manford, M. (2020). A review of seizures and epilepsy following traumatic brain injury [Review]. *Journal of Neurology*, *267*(10), 3105-3111. https://doi.org/10.1007/s00415-020-09926-w

Forslund, M. V., Perrin, P. B., Roe, C., Sigurdardottir, S., Hellstrom, T., Berntsen, S. A., Lu, J., Arango-Lasprilla, J. C., & Andelic, N. (2019). Global Outcome Trajectories up to 10 Years After Moderate to Severe Traumatic Brain Injury. *Frontiers in Neurology*, *10*, Article 219. https://doi.org/10.3389/fneur.2019.00219

Fraser, A., Macdonald-Wallis, C., Tilling, K., Boyd, A., Golding, J., Davey Smith, G., Henderson, J., Macleod, J., Molloy, L., & Ness, A. (2013). Cohort profile: the Avon Longitudinal Study of Parents and Children: ALSPAC mothers cohort. *International journal of epidemiology*, *42*(1), 97-110. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3600619/pdf/dys066.pdf

Freeman, C. R., Wiers, C. E., Sloan, M. E., Zehra, A., Ramirez, V., Wang, G. J., & Volkow, N. D. (2018). Emotion Recognition Biases in Alcohol Use Disorder. *Alcoholism-*

Clinical and Experimental Research, *42*(8), 1541-1547. https://doi.org/10.1111/acer.13802

Friedland, D. P. (2013). Improving the Classification of Traumatic Brain Injury: The Mayo Classification System for Traumatic Brain Injury Severity [Review Article]. *Journal of Spine*, *S4: 005*. https://doi.org/10.4172/2165-7939.S4-005

Friesen, W. V., & Ekman, P. (1976). *Pictures of facial affect*. Consulting psychologists press.

Frost, R. B., Farrer, T. J., Primosch, M., & Hedges, D. W. (2013). Prevalence of Traumatic Brain Injury in the General Adult Population: A Meta-Analysis. *Neuroepidemiology*, *40*(3), 154-159. https://doi.org/10.1159/000343275

Funkiewiez, A., Bertoux, M., de Souza, L. C., Levy, R., & Dubois, B. (2012). The SEA (Social Cognition and Emotional Assessment): A Clinical Neuropsychological Tool for Early Diagnosis of Frontal Variant of Frontotemporal Lobar Degeneration [Article]. *Neuropsychology*, *26*(1), 81-90. https://doi.org/10.1037/a0025318

Fure, S. C. R., Howe, E. I., Spjelkavik, O., Roe, C., Rike, P. O., Olsen, A., Ponsford, J., Andelic, N., & Lovstad, M. (2021). Post-concussion symptoms three months after mild-to-moderate TBI: characteristics of sick-listed patients referred to specialized treatment and consequences of intracranial injury. *Brain Injury*, *35*(9), 1054-1064. https://doi.org/10.1080/02699052.2021.1953593

Galgano, M., Toshkezi, G., Qiu, X. C., Russell, T., Chin, L., & Zhao, L. R. (2017). Traumatic Brain Injury: Current Treatment Strategies and Future Endeavors [Review]. *Cell Transplantation*, *26*(7), 1118-1130. https://doi.org/10.1177/0963689717714102

Gasquoine, P. G. (2020). Historical perspectives on evolving operational definitions of concussive brain injury: From railway spine to sport-related concussion [Review]. *Clinical Neuropsychologist*, *34*(2), 278-295. https://doi.org/10.1080/13854046.2019.1621383

Giza, C. C., & Prins, M. L. (2006). Is being plastic fantastic? Mechanisms of altered plasticity after developmental traumatic brain injury. *Developmental Neuroscience*, *28*(4-5), 364-379. https://doi.org/10.1159/000094163

Goeleven, E., De Raedt, R., Leyman, L., & Verschuere, B. (2008). The Karolinska Directed Emotional Faces: A validation study. *Cognition & Emotion*, *22*(6), 1094-1118. https://doi.org/10.1080/02699930701626582

Gould, K. R., Ponsford, J., Johnston, L., & Schonberger, M. (2011). The nature, frequency and course of psychiatric disorders in the first year after traumatic brain injury: a prospective study. *Psychological Medicine*, *41*(10), 2099-2109. https://doi.org/10.1017/s003329171100033x

Grier, J. B. (1971). NONPARAMETRIC INDEXES FOR SENSITIVITY AND BIAS - COMPUTING FORMULAS [Article]. *Psychological Bulletin*, *75*(6), 424-&. https://doi.org/10.1037/h0031246

Griffiths, S., Jarrold, C., Penton-Voak, I. S., Woods, A. T., Skinner, A. L., & Munafo, M. R. (2019). Impaired Recognition of Basic Emotions from Facial Expressions in Young People with Autism Spectrum Disorder: Assessing the Importance of Expression Intensity. *Journal of Autism and Developmental Disorders*, *49*(7), 2768-2778. https://doi.org/10.1007/s10803-017-3091-7

Grimes, D. A., & Schulz, K. F. (2002). Cohort studies: marching towards outcomes. *Lancet*, *359*(9303), 341-345. https://doi.org/10.1016/s0140-6736(02)07500-1

Harrell, F. (2014). *Hmisc: A package of miscellaneous R functions*. In Programs available from https://hbiostat.org/R/Hmisc/

Harris, P. A., Taylor, R., Thielke, R., Payne, J., Gonzalez, N., & Conde, J. G. (2009). Research electronic data capture (REDCap)-A metadata-driven methodology and workflow process for providing translational research informatics support

[Article]. *Journal of Biomedical Informatics*, *42*(2), 377-381. https://doi.org/10.1016/j.jbi.2008.08.010
Harron, K., Doidge, J. C., & Goldstein, H. (2020). Assessing data linkage quality in cohort studies. *Annals of Human Biology*, *47*(2), 218-226.
Haxby, J. V., & Gobbini, M. I. (2011). *Distributed neural systems for face perception*. The Oxford Handbook of Face Perception.
Hayes, G. S., McLennan, S. N., Henry, J. D., Phillips, L. H., Terrett, G., Rendell, P. G., Pelly, R. M., & Labuschagne, I. (2020). Task Characteristics Influence Facial Emotion Recognition Age-Effects: A Meta-Analytic Review. *Psychology and Aging*, *35*(2), 295-315. https://doi.org/10.1037/pag0000441
Headway. (2018). *Acquired Brain Injury: The numbers behind the hidden disability*. Headway - the brain injury association. Retrieved 4th January from
Hebbali, A. (2017). *olsrr - Tools for Building OLS Regression Models*. In (Version 0.5.3)
Hedge, C., Powell, G., & Sumner, P. (2018). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences [Article]. *Behavior Research Methods*, *50*(3), 1166-1186. https://doi.org/10.3758/s13428-017-0935-1
Held, J., Visla, A., Zinbarg, R. E., Wolfer, C., & Fluckiger, C. (2020). How do worry and clinical status impact working memory performance? An experimental investigation. *BMC psychiatry*, *20*(1). https://doi.org/10.1186/s12888-020-02694-x
Henry, A., Bakchine, S., Maarouf, A., Chaunu, M. P., Rumbach, L., Magnin, E., Tourbah, A., & Montreuil, M. (2015). Facial Emotion Recognition and Faux Pas Interpretation in Multiple Sclerosis. *Brain Impairment*, *16*(3), 158-172. https://doi.org/10.1017/BrImp.2015.33
Henry, J. D., Cowan, D. G., Lee, T., & Sachdev, P. S. (2015). Recent trends in testing social cognition. *Current Opinion in Psychiatry*, *28*(2), 133-140. https://doi.org/10.1097/yco.0000000000000139
Henry, J. D., Phillips, L. H., Crawford, J. R., Theodorou, G., & Summers, F. (2006). Cognitive and psychosocial correlates of alexithymia following traumatic brain injury [Article]. *Neuropsychologia*, *44*(1), 62-72. https://doi.org/10.1016/j.neuropsychologia.2005.04.011
Henry, J. D., von Hippel, W., Molenberghs, P., Lee, T., & Sachdev, P. S. (2016). Clinical assessment of social cognitive function in neurological disorders [Review]. *Nature Reviews Neurology*, *12*(1), 12. https://doi.org/10.1038/nrneurol.2015.229
Hicks, A. J., Gould, K. R., Hopwood, M., Kenardy, J., Krivonos, I., & Ponsford, J. (2017). Behaviours of concern following moderate to severe traumatic brain injury in individuals living in the community. *Brain Injury*, *31*(10), 1312-1319. https://doi.org/10.1080/02699052.2017.1317361
Hilbert, S., Nakagawa, T. T., Puci, P., Zech, A., & Buhner, M. (2015). The Digit Span Backwards Task Verbal and Visual Cognitive Strategies in Working Memory Assessment [Article]. *European Journal of Psychological Assessment*, *31*(3), 174-180. https://doi.org/10.1027/1015-5759/a000223
Hoaken, P. N., Allaby, D. B., & Earle, J. (2007). Executive cognitive functioning and the recognition of facial expressions of emotion in incarcerated violent offenders, non-violent offenders, and controls. *Aggress Behav*, *33*(5), 412-421. https://doi.org/10.1002/ab.20194
Hockey, A., & Geffen, G. (2004). The concurrent validity and test-retest reliability of a visuospatial working memory task. *Intelligence*, *32*(6), 591-605. https://doi.org/10.1016/j.intell.2004.07.009

Hoffmann, H., Kessler, H., Eppel, T., Rukavina, S., & Traue, H. C. (2010). Expression intensity, gender and facial emotion recognition: Women recognize only subtle facial emotions better than men. *Acta Psychologica*, *135*(3), 278-283. https://doi.org/10.1016/j.actpsy.2010.07.012

Hogan, T. P., Benjamin, A., & Brezinski, K. L. (2000). Reliability methods: A note on the frequency of use of various types. *Educational and psychological measurement*, *60*(4), 523-531.

Honan, C. A., McDonald, S., Sufani, C., Hine, D. W., & Kumfor, F. (2016). The awareness of social inference test: development of a shortened version for use in adults with acquired brain injury [Article]. *Clinical Neuropsychologist*, *30*(2), 243-264. https://doi.org/10.1080/13854046.2015.1136691

Honan, C. A., McDonald, S., Tate, R., Ownsworth, T., Togher, L., Fleming, J., Anderson, V., Morgan, A., Catroppa, C., Douglas, J., Francis, H., Wearne, T., Sigmundsdottir, L., & Ponsford, J. (2019). Outcome instruments in moderate-to-severe adult traumatic brain injury: recommendations for use in psychosocial research. *Neuropsychological Rehabilitation*, *29*(6), 896-916. https://doi.org/10.1080/09602011.2017.1339616

Hopkins, M. J., Dywan, J., & Segalowitz, S. J. (2002). Altered electrodermal response to facial expression after closed head injury. *Brain Injury*, *16*(3), 245-257. https://doi.org/10.1080/02699050110103346

Horning, S. M., Cornwell, R. E., & Davis, H. P. (2012). The recognition of facial expressions: an investigation of the influence of age and cognition. *Neuropsychol Dev Cogn B Aging Neuropsychol Cogn*, *19*(6), 657-676. https://doi.org/10.1080/13825585.2011.645011

Howieson, D. (2019). Current limitations of neuropsychological tests and assessment procedures [Article]. *Clinical Neuropsychologist*, *33*(2), 200-208. https://doi.org/10.1080/13854046.2018.1552762

Ietswaart, M., Milders, M., Crawford, J. R., Currie, D., & Scott, C. L. (2008). Longitudinal aspects of emotion recognition in patients with traumatic brain injury. *Neuropsychologia*, *46*(1), 148-159. https://doi.org/10.1016/j.neuropsychologia.2007.08.002

Ihme, K., Sacher, J., Lichev, V., Rosenberg, N., Kugel, H., Rufer, M., Grabe, H. J., Pampel, A., Lepsien, J., Kersting, A., Villringer, A., Lane, R. D., & Suslow, T. (2014). Alexithymic features and the labeling of brief emotional facial expressions - An fMRI study. *Neuropsychologia*, *64*, 289-299. https://doi.org/10.1016/j.neuropsychologia.2014.09.044

Ihme, K., Sacher, J., Lichev, V., Rosenberg, N., Kugel, H., Rufer, M., Grabe, H. J., Pampel, A., Lepsien, J., Kersting, A., Villringer, A., & Suslow, T. (2014). Alexithymia and the labeling of facial emotions: response slowing and increased motor and somatosensory processing [Article]. *Bmc Neuroscience*, *15*, 10, Article 40. https://doi.org/10.1186/1471-2202-15-40

Israelashvili, J., Pauw, L. S., Sauter, D. A., & Fischer, A. H. (2021). Emotion Recognition from Realistic Dynamic Emotional Expressions Cohere with Established Emotion Recognition Tests: A Proof-of-Concept Validation of the Emotional Accuracy Test. *Journal of Intelligence*, *9*(2), Article 25. https://doi.org/10.3390/jintelligence9020025

Jack, R. E., Garrod, O. G. B., & Schyns, P. G. (2014). Dynamic Facial Expressions of Emotion Transmit an Evolving Hierarchy of Signals over Time. *Current Biology*, *24*(2), 187-192. https://doi.org/10.1016/j.cub.2013.11.064

Jack, R. E., Sun, W., Delis, I., Garrod, O. G. B., & Schyns, P. G. (2016). Four Not Six: Revealing Culturally Common Facial Expressions of Emotion. *Journal of*

Experimental Psychology-General, *145*(6), 708-730. https://doi.org/10.1037/xge0000162

Jaeggi, S. M., Buschkuehl, M., Perrig, W. J., & Meier, B. (2010). The concurrent validity of the N-back task as a working memory measure. *Memory*, *18*(4), 394-412, Article Pii 921420785. https://doi.org/10.1080/09658211003702171

Jorge, R. E., Robinson, R. G., Moser, D., Tateno, A., Crespo-Facorro, B., & Arndt, S. (2004). Major depression following traumatic brain injury. *Archives of General Psychiatry*, *61*(1), 42-50. https://doi.org/10.1001/archpsyc.61.1.42

Jorna, L. S., Westerhof-Evers, H. J., Khosdelazad, S., Rakers, S. E., van der Naalt, J., Groen, R. J. M., Buunk, A. M., & Spikman, J. M. (2021). Behaviors of Concern after Acquired Brain Injury: The Role of Negative Emotion Recognition and Anger Misattribution. *Journal of the International Neuropsychological Society*, *27*(10), 1015-1023, Article Pii s135561772000140x. https://doi.org/10.1017/s135561772000140x

Joseph, B., Khan, M., & Rhee, P. (2018). Non-invasive diagnosis and treatment strategies for traumatic brain injury: an update [Review]. *Journal of Neuroscience Research*, *96*(4), 589-600. https://doi.org/10.1002/jnr.24132

Joseph, D. L., & Newman, D. A. (2010). Emotional Intelligence: An Integrative Meta-Analysis and Cascading Model. *Journal of Applied Psychology*, *95*(1), 54-78. https://doi.org/10.1037/a0017286

Jourdan, C., Azouvi, P., Genet, F., Selly, N., Josseran, L., & Schnitzler, A. (2018). Disability and Health Consequences of Traumatic Brain Injury National Prevalence [Article]. *American Journal of Physical Medicine & Rehabilitation*, *97*(5), 323-331. https://doi.org/10.1097/phm.0000000000000848

Jourdan, C., Bayen, E., Pradat-Diehl, P., Ghout, I., Darnoux, E., Azerad, S., Vallat-Azouvi, C., Charanton, J., Aegerter, P., Ruet, A., & Azouvi, P. (2016). A comprehensive picture of 4-year outcome of severe brain injuries. Results from the PariS-TBI study. *Annals of Physical and Rehabilitation Medicine*, *59*(2), 100-106. https://doi.org/10.1016/j.rehab.2015.10.009

Kelly, M., McDonald, S., & Frith, M. H. J. (2017). A Survey of Clinicians Working in Brain Injury Rehabilitation: Are Social Cognition Impairments on the Radar? *Journal of Head Trauma Rehabilitation*, *32*(4), E55-E65. https://doi.org/10.1097/htr.0000000000000269

Keltner, D., Sauter, D., Tracy, J., & Cowen, A. (2019). Emotional Expression: Advances in Basic Emotion Theory. *Journal of Nonverbal Behavior*, *43*(2), 133-160. https://doi.org/10.1007/s10919-019-00293-3

Keltner, D., Tracy, J. L., Sauter, D., & Cowen, A. (2019). What Basic Emotion Theory Really Says for the Twenty-First Century Study of Emotion. *Journal of Nonverbal Behavior*, *43*(2), 195-201. https://doi.org/10.1007/s10919-019-00298-y

Kennedy, E., Heron, J., & Munafo, M. (2017). Substance use, criminal behaviour and psychiatric symptoms following childhood traumatic brain injury: findings from the ALSPAC cohort. *European Child & Adolescent Psychiatry*, *26*(10), 1197-1206. https://doi.org/10.1007/s00787-017-0975-1

Kessels, R. P., Montagne, B., Hendriks, A. W., Perrett, D. I., & de Haan, E. H. (2014). Assessment of perception of morphed facial expressions using the Emotion Recognition Task: normative data from healthy participants aged 8-75. *J Neuropsychol*, *8*(1), 75-93. https://doi.org/10.1111/jnp.12009

Kessels, R. P. C. (2019). Improving precision in neuropsychological assessment: Bridging the gap between classic paper-and-pencil tests and paradigms from cognitive neuroscience [Review]. *Clinical Neuropsychologist*, *33*(2), 357-368. https://doi.org/10.1080/13854046.2018.1518489

Khosdelazad, S., Jorna, L. S., McDonald, S., Rakers, S. E., Huitema, R. B., Buunk, A. M., & Spikman, J. M. (2020). Comparing static and dynamic emotion recognition tests: Performance of healthy participants. *PLoS One*, *15*(10), Article e0241297. https://doi.org/10.1371/journal.pone.0241297

Kirchner, W. K. (1958). Age differences in short-term retention of rapidly changing information. *Journal of experimental psychology*, *55*(4), 352.

Knapp, M. L., Hall, J. A., & Horgan, T. G. (2013). *Nonverbal communication in human interaction*. Cengage Learning.

Knowles, K. A., & Olatunji, B. O. (2020). Specificity of trait anxiety in anxiety and depression: Meta-analysis of the State-Trait Anxiety Inventory. *Clinical Psychology Review*, *82*, Article 101928. https://doi.org/10.1016/j.cpr.2020.101928

Knox, L., & Douglas, J. (2009). Long-term ability to interpret facial expression after traumatic brain injury and its relation to social integration. *Brain and Cognition*, *69*(2), 442-449. https://doi.org/10.1016/j.bandc.2008.09.009

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of chiropractic medicine*, *15*(2), 155-163.

Krause, F. C., Linardatos, E., Fresco, D. M., & Moore, M. T. (2021). Facial emotion recognition in major depressive disorder: A meta-analytic review. *Journal of affective disorders*, *293*, 320-328. https://doi.org/10.1016/j.jad.2021.06.053

Krumhuber, E. G., Kappas, A., & Manstead, A. S. R. (2013). Effects of Dynamic Aspects of Facial Expressions: A Review [Review]. *Emotion Review*, *5*(1), 41-46. https://doi.org/10.1177/1754073912451349

Kwasnicka, D., Keller, J., Perski, O., Potthoff, S., Ten Hoor, G. A., Ainsworth, B., Crutzen, R., Dohle, S., Van Dongen, A., & Heino, M. (2022). White Paper: Open Digital Health–accelerating transparent and scalable health promotion and treatment. *Health Psychology Review*, 1-17.

Lakens, D. (2013a, 2019). *Calculating and Reporting Effect Sizes to Facilitate Cumulative Science: A Practical Primer for t-tests and ANOVAs*. Retrieved from osf.io/ixgcd

Lakens, D. (2013b). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, *4*, 863. <Go to WoS>://WOS:000405075500001

Lakens, D. (2017). Equivalence Tests: A Practical Primer for t Tests, Correlations, and Meta-Analyses [Article]. *Social Psychological and Personality Science*, *8*(4), 355-362. https://doi.org/10.1177/1948550617697177

Lakens, D., Adolfi, F. G., Albers, C. J., Anvari, F., Apps, M. A. J., Argamon, S. E., Baguley, T., Becker, R. B., Benning, S. D., Bradford, D. E., Buchanan, E. M., Caldwell, A. R., Van Calster, B., Carlsson, R., Chen, S. C., Chung, B., Colling, L. J., Collins, G. S., Crook, Z., Cross, E. S., Daniels, S., Danielsson, H., DeBruine, L., Dunleavy, D. J., Earp, B. D., Feist, M. I., Ferrell, J. D., Field, J. G., Fox, N. W., Friesen, A., Gomes, C., Gonzalez-Marquez, M., Grange, J. A., Grieve, A. P., Guggenberger, R., Grist, J., van Harmelen, A. L., Hasselman, F., Hochard, K. D., Hoffarth, M. R., Holmes, N. P., Ingre, M., Isager, P. M., Isotalus, H. K., Johansson, C., Juszczyk, K., Kenny, D. A., Khalil, A. A., Konat, B., Lao, J. P., Larsen, E. G., Lodder, G. M. A., Lukavsky, J., Madan, C. R., Manheim, D., Martin, S. R., Martin, A. E., Mayo, D. G., McCarthy, R. J., McConway, K., McFarland, C., Nio, A. Q. X., Nilsonne, G., de Oliveira, C. L., de Xivry, J. J. O., Parsons, S., Pfuhl, G., Quinn, K. A., Sakon, J. J., Saribay, S. A., Schneider, I. K., Selvaraju, M., Sjoerds, Z., Smith, S. G., Smits, T., Spies, J. R., Sreekumar, V., Steltenpohl, C. N., Stenhouse, N., Swiatkowski, W., Vadillo, M. A., Van Assen, M., Williams, M. N., Williams, S. E., Williams, D. R., Yarkoni, T., Ziano,

I., & Zwaan, R. A. (2018). Justify your alpha. *Nature Human Behaviour*, *2*(3), 168-171. https://doi.org/10.1038/s41562-018-0311-x

Lakens, D., McLatchie, N., Isager, P. M., Scheel, A. M., & Dienes, Z. (2020). Improving Inferences About Null Effects With Bayes Factors and Equivalence Tests. *Journals of Gerontology Series B-Psychological Sciences and Social Sciences*, *75*(1), 45-57. https://doi.org/10.1093/geronb/gby065

Lawrence, K., Campbell, R., & Skuse, D. (2015). Age, gender, and puberty influence the development of facial emotion recognition. *Front Psychol*, *6*, 761. https://doi.org/10.3389/fpsyg.2015.00761

Lee, I. A., & Preacher, K. J. (2013). *Calculation for the test of the difference between two dependent correlations with one variable in common*. In http://quantpsy.org

Lesser, I. M. (1981). A REVIEW OF THE ALEXITHYMIA CONCEPT [Review]. *Psychosomatic Medicine*, *43*(6), 531-543. https://doi.org/10.1097/00006842-198112000-00009

Levin, H. S., & Diaz-Arrastia, R. R. (2015). Diagnosis, prognosis, and clinical management of mild traumatic brain injury [Review]. *Lancet Neurology*, *14*(5), 506-517. https://doi.org/10.1016/s1474-4422(15)00002-2

Lewis, G., Pelosi, A. J., Araya, R., & Dunn, G. (1992). Measuring psychiatric disorder in the community: a standardized assessment for use by lay interviewers. *Psychological Medicine*, *22*(2), 465-486. https://doi.org/10.1017/S0033291700030415

Lexell, J. E., & Downham, D. Y. (2005). How to assess the reliability of measurements in rehabilitation.

Lin, X. G., Zhang, X. L., Liu, Q. Q., Zhao, P. W., Zhang, H., Wang, H. S., & Yi, Z. Q. (2021). Theory of mind in adults with traumatic brain injury: A meta-analysis. *Neuroscience and Biobehavioral Reviews*, *121*, 106-118. https://doi.org/10.1016/j.neubiorev.2020.12.010

LLC, S. *Stata*. In (Version 16) Stata Press.

Logan, G. D., & Cowan, W. B. (1984). On the ability to inhibit thought and action: A theory of an act of control. *Psychological review*, *91*(3), 295.

Lovibond, P. F., & Lovibond, S. H. (1995). THE STRUCTURE OF NEGATIVE EMOTIONAL STATES - COMPARISON OF THE DEPRESSION ANXIETY STRESS SCALES (DASS) WITH THE BECK DEPRESSION AND ANXIETY INVENTORIES [Article]. *Behaviour Research and Therapy*, *33*(3), 335-343. https://doi.org/10.1016/0005-7967(94)00075-u

Maas, A. I. R., Menon, D. K., Adelson, P. D., Andelic, N., Bell, M. J., Belli, A., Bragge, P., Brazinova, A., Burki, A., Chesnut, R. M., Citerio, G., Coburn, M., Cooper, D. J., Crowder, A. T., Czeiter, E., Czosnyka, M., Diaz-Arrastia, R., Dreier, J. P., Duhaime, A. C., Ercole, A., van Essen, T. A., Feigin, V. L., Gao, G. Y., Giacino, J., Gonzalez-Lara, L. E., Gruen, R. L., Gupta, D., Hartings, J. A., Hill, S., Jiang, J. Y., Ketharanathan, N., Kompanje, E. J. O., Lanyon, L., Laureys, S., Lecky, F., Levin, H., Lingsma, H. F., Maegele, M., Majdan, M., Manley, G., Marsteller, J., Mascia, L., McFadyen, C., Mondello, S., Newcombe, V., Palotie, A., Parizel, P. M., Peul, W., Piercy, J., Polinder, S., Puybasset, L., Rasmussen, T. E., Rossaint, R., Smielewski, P., Soderberg, J., Stanworth, S. J., Stein, M. B., von Steinbuchel, N., Stewart, W., Steyerberg, E. W., Stocchetti, N., Synnot, A., Ao, B. T., Tenovuo, O., Theadom, A., Tibboel, D., Videtta, W., Wang, K. K. W., Williams, W. H., Wilson, L., Yaffe, K., & In, T. P. I. (2017). Traumatic brain injury: integrated approaches to improve prevention, clinical care, and research. *Lancet Neurology*, *16*(12), 987-1048. https://doi.org/10.1016/s1474-4422(17)30371-x

Maas, A. I. R., Stocchetti, N., & Bullock, R. (2008). Moderate and severe traumatic brain injury in adults. *The Lancet Neurology*, *7*(8), 728-741. https://doi.org/https://doi.org/10.1016/S1474-4422(08)70164-9

Maggio, M. G., Maresca, G., Stagnitti, M. C., Anchesi, S., Casella, C., Pajno, V., De Luca, R., Manuli, A., & Calabrò, R. S. (2020). Social cognition in patients with acquired brain lesions: An overview on an under-reported problem. *Applied Neuropsychology: Adult*, 1-13. https://doi.org/10.1080/23279095.2020.1753058

Mahedy, L., Wootton, R. E., Suddell, S., Skirrow, C., Field, M., Heron, J., Hickman, M., & Munafo, M. (2020). Testing the association between tobacco and cannabis use and cognitive functioning: Findings from an observational and Mendelian randomization study. *medRxiv*.

Majdan, M., Plancikova, D., Brazinova, A., Rusnak, M., Nieboer, D., Feigin, V., & Maas, A. (2016). Epidemiology of traumatic brain injuries in Europe: a cross-sectional analysis. *Lancet Public Health*, *1*(2), E76-E83. https://doi.org/10.1016/s2468-2667(16)30017-2

Malec, J. F., Brown, A. W., Leibson, C. L., Flaada, J. T., Mandrekar, J. N., Diehl, N. N., & Perkins, P. K. (2007). The Mayo classification system for traumatic brain injury severity [Article]. *Journal of Neurotrauma*, *24*(9), 1417-1424. https://doi.org/10.1089/neu.2006.0245

Manierka, M. S., Rezaei, R., Palacios, S., Haigh, S. M., & Hutsler, J. J. (2021). In the Mood to Be Social: Affective State Influences Facial Emotion Recognition in Healthy Adults. *Emotion*, *21*(7), 1576-1581. https://doi.org/10.1037/emo0000999

Mann, C. J. (2003). Observational research methods. Research design II: cohort, cross sectional, and case-control studies. *Emergency Medicine Journal*, *20*(1), 54-60. https://doi.org/10.1136/emj.20.1.54

Maoz, K., Eldar, S., Stoddard, J., Pine, D. S., Leibenluft, E., & Bar-Haim, Y. (2016). Angry-happy interpretations of ambiguous faces in social anxiety disorder. *Psychiatry Res*, *241*, 122-127. https://doi.org/10.1016/j.psychres.2016.04.100

Martinez-Sanchez, F., Fernandez-Abascal, E. G., & Sanchez-Perez, N. (2017). Recognition of Emotional Facial Expressions in Alexithymia [Article]. *Studia Psychologica*, *59*(3), 206-216. https://doi.org/10.21909/sp.2017.03.741

Massonnie, J. (2019). *Working Memory - Backward Digit Span*. In

Mathias, J. L., & Wheaton, P. (2007). Changes in attention and information-processing speed following severe traumatic brain injury: A meta-analytic review. *Neuropsychology*, *21*(2), 212-223. https://doi.org/10.1037/0894-4105.21.2.212

May, M., Milders, M., Downey, B., Whyte, M., Higgins, V., Wojcik, Z., Amin, S., & O'Rourke, S. (2017). Social Behavior and Impairments in Social Cognition Following Traumatic Brain Injury. *Journal of the International Neuropsychological Society*, *23*(5), 400-411. https://doi.org/10.1017/s1355617717000182

McAllister, T. W. (2011). Neurobiological consequences of traumatic brain injury. *Dialogues in clinical neuroscience*, *13*(3), 287-300. https://doi.org/10.31887/DCNS.2011.13.2/tmcallister

McDonald, S. (2013). Impairments in Social Cognition Following Severe Traumatic Brain Injury. *Journal of the International Neuropsychological Society*, *19*(3), 231-246. https://doi.org/10.1017/s1355617712001506

McDonald, S., Bornhofen, C., Shum, D., Long, E., Saunders, C., & Neulinger, K. (2006). Reliability and validity of The Awareness of Social Inference Test (TASIT): A clinical test of social perception [Review]. *Disability and rehabilitation*, *28*(24), 1529-1542. https://doi.org/10.1080/09638280600646185

McDonald, S., Flanagan, S., Martin, I., & Saunders, C. (2004). The ecological validity of TASIT: A test of social perception [Article]. *Neuropsychological Rehabilitation*, *14*(3), 285-302. https://doi.org/10.1080/09602010343000237

McDonald, S., Flanagar, S., Rollins, J., & Kinch, J. (2003). A new clinical tool for assessing social perception after traumatic brain injury [Article]. *Journal of Head Trauma Rehabilitation*, *18*(3), 219-238. https://doi.org/10.1097/00001199-200305000-00001

McDonald, S., Honan, C., Allen, S. K., El-Helou, R., Kelly, M., Kumfor, F., Piguet, O., Hazelton, J. L., Padgett, C., & Keage, H. A. D. (2018). Normal adult and adolescent performance on TASIT-S, a short version of The Assessment of Social Inference Test [Article]. *Clinical Neuropsychologist*, *32*(4), 700-719. https://doi.org/10.1080/13854046.2017.1400106

McDonald, S., Rosenfeld, J., Henry, J. D., Togher, L., Tate, R., & Bornhofen, C. (2011). Emotion Perception and Alexithymia in People With Severe Traumatic Brain Injury: One Disorder or Two? A Preliminary Investigation. *Brain Impairment*, *12*(3), 165-178. <Go to ISI>://WOS:000299912000001

McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological methods*, *1*(1), 30.

McKee, A. C., & Daneshvar, D. H. (2015). The neuropathology of traumatic brain injury. *Handbook of clinical neurology*, *127*, 45-66. https://doi.org/10.1016/B978-0-444-52892-6.00004-0

McLean, C. P., Asnaani, A., Litz, B. T., & Hofmann, S. G. (2011). Gender differences in anxiety disorders: Prevalence, course of illness, comorbidity and burden of illness. *Journal of Psychiatric Research*, *45*(8), 1027-1035. https://doi.org/10.1016/j.jpsychires.2011.03.006

McLellan, T., & McKinlay, A. (2013). Sensitivity to emotion, empathy and theory of mind: Adult performance following childhood TBI. *Brain Injury*, *27*(9), 1032-1037. https://doi.org/10.3109/02699052.2013.794965

Megreya, A. M., & Burton, A. M. (2007). Hits and false positives in face matching: A familiarity-based dissociation. *Perception & Psychophysics*, *69*(7), 1175-1184. https://doi.org/10.3758/bf03193954

Mendes Ferrer Rosa, N., Ferreira Borges, V., Cheffer, L., Torro Alves, N., & Estanislau, C. (2017). Symptoms of Anxiety and Depression Modulate the Recognition of Facial Emotion. *Universitas Psychologica*, *16*(4). https://doi.org/10.11144/Javeriana.upsy16-4.sadm

Menon, D. K., Schwab, K., Wright, D. W., & Maas, A. I. (2010). Position Statement: Definition of Traumatic Brain Injury [Article]. *Archives of Physical Medicine and Rehabilitation*, *91*(11), 1637-1640. https://doi.org/10.1016/j.apmr.2010.05.017

Messick, S. (1980). TEST VALIDITY AND THE ETHICS OF ASSESSMENT [Article]. *American psychologist*, *35*(11), 1012-1027. https://doi.org/10.1037/0003-066x.35.11.1012

Meyer, G. J., Finn, S. E., Eyde, L. D., Kay, G. G., Moreland, K. L., Dies, R. R., Eisman, E. J., Kubiszyn, T. W., & Reed, G. M. (2001). Psychological testing and psychological assessment: A review of evidence and issues. *American psychologist*, *56*(2), 128.

Mikolic, A., van Klaveren, D., Groeniger, J. O., Wiegers, E. J. A., Lingsma, H. F., Zeldovich, M., von Steinbuchel, N., Maas, A. I. R., van Lennep, J. R. E., Polinder, S., & Participants, C.-T. (2021). Differences between Men and Women in Treatment and Outcome after Traumatic Brain Injury. *Journal of Neurotrauma*, *38*(2), 235-251. https://doi.org/10.1089/neu.2020.7228

Milders, M. (2019). Relationship between social cognition and social behaviour following traumatic brain injury [Article]. *Brain Injury*, *33*(1), 62-68. https://doi.org/10.1080/02699052.2018.1531301

Milders, M., Fuchs, S., & Crawford, J. R. (2003). Neuropsychological impairments and changes in emotional and social behaviour following severe traumatic brain injury. *Journal of Clinical and Experimental Neuropsychology*, *25*(2), 157-172. https://doi.org/10.1076/jcen.25.2.157.13642

Milders, M., Ietswaart, M., Crawford, J. R., & Currie, D. (2008). Social behavior following traumatic brain injury and its association with emotion recognition, understanding of intentions, and cognitive flexibility [Article]. *Journal of the International Neuropsychological Society*, *14*(2), 318-326. https://doi.org/10.1017/s1355617708080351

Mill, A., Allik, J., Realo, A., & Valk, R. (2009). Age-Related Differences in Emotion Recognition Ability: A Cross-Sectional Study. *Emotion*, *9*(5), 619-630. https://doi.org/10.1037/a0016562

Mitrushina, M., Boone, K. B., Razani, J., & D'Elia, L. F. (2005). *Handbook of normative data for neuropsychological assessment*. Oxford University Press.

Montagne, B., Kessels, R. P. C., De Haan, E. H. F., & Perrett, D. I. (2007). The emotion recognition task: A paradigm to measure the perception of facial emotional expressions at different intensities. *Perceptual and Motor Skills*, *104*(2), 589-598. https://doi.org/10.2466/pms.104.2.589-598

Müller-Glodde, M. (2015). *Emotion Perception in Brain Injury Patients: Assessing the Feasibility of the Bristol Emotion Recognition Task* University of Bristol]. [Unpublished master's thesis].

Munafo, M. R., Tilling, K., Taylor, A. E., Evans, D. M., & Smith, G. D. (2018). Collider scope: when selection bias can substantially influence observed associations. *International journal of epidemiology*, *47*(1), 226-235. https://doi.org/10.1093/ije/dyx206

Murphy, J., Millgate, E., Geary, H., Catmur, C., & Bird, G. (2019). No effect of age on emotion recognition after accounting for cognitive factors and depression [Article]. *Quarterly Journal of Experimental Psychology*, *72*(11), 2690-2704. https://doi.org/10.1177/1747021819859514

Murphy, J. M., Bennett, J. M., de la Piedad Garcia, X., & Willis, M. L. (2021). Emotion Recognition and Traumatic Brain Injury: a Systematic Review and Meta-Analysis. *Neuropsychology Review*. https://doi.org/10.1007/s11065-021-09510-7

Mushkudiani, N. A., Engel, D. C., Steyerberg, E. W., Butcher, I., Lu, J., Marmarou, A., Slieker, F., McHugh, G. S., Murray, G. D., & Maas, A. I. R. (2007). Prognostic value of demographic characteristics in traumatic brain injury: Results from the IMPACT study. *Journal of Neurotrauma*, *24*(2), 259-269. https://doi.org/10.1089/neu.2006.0028

NAMHC. (2016). *Behavioral Assessment Methods for RDoC Constructs: A Report by the National Advisory Mental Health Council Workgroup on Tasks and Measures for Research Domain Criteria (RDoC)* Retrieved 04/11/2019 from https://www.nimh.nih.gov/about/advisory-boards-and-groups/namhc/reports/behavioral-assessment-methods-for-rdoc-constructs.shtml

Neumann, D., Keiski, M. A., McDonald, B. C., & Wang, Y. (2014). Neuroimaging and facial affect processing: implications for traumatic brain injury. *Brain Imaging Behav*, *8*(3), 460-473. https://doi.org/10.1007/s11682-013-9285-5

Neumann, D., Malec, J. F., & Hammond, F. M. (2017). Negative Attribution Bias and Anger After Traumatic Brain Injury [Article]. *Journal of Head Trauma Rehabilitation*, *32*(3), 197-204. https://doi.org/10.1097/htr.0000000000000259

Neumann, D., Sander, A. M., Perkins, S. M., Bhamidipalli, S. S., & Hammond, F. M. (2021). Negative Attribution Bias and Related Risk Factors After Brain Injury

[Article]. *Journal of Head Trauma Rehabilitation*, *36*(1), E61-E70. https://doi.org/10.1097/htr.0000000000000600

Neumann, D., Zupan, B., Malec, J. F., & Hammond, F. (2014). Relationships Between Alexithymia, Affect Recognition, and Empathy After Traumatic Brain Injury [Article]. *Journal of Head Trauma Rehabilitation*, *29*(1), E18-E27. https://doi.org/10.1097/HTR.0b013e31827fb0b5

Neumann, R., Volker, J., Hajba, Z., & Seiler, S. (2021). Lesions and reduced working memory impair emotion recognition in self and others. *Cognition & Emotion*, *35*(8), 1527-1542. https://doi.org/10.1080/02699931.2021.1983521

Newen, A., Welpinghus, A., & Juckel, G. (2015). Emotion Recognition as Pattern Recognition: The Relevance of Perception. *Mind & Language*, *30*(2), 187-208. https://doi.org/10.1111/mila.12077

Newton, P. E., & Shaw, S. D. (2013). Standards for Talking and Thinking About Validity [Article]. *Psychological methods*, *18*(3), 301-319. https://doi.org/10.1037/a0032969

Ng, S. Y., & Lee, A. Y. W. (2019). Traumatic Brain Injuries: Pathophysiology and Potential Therapeutic Targets [Review]. *Frontiers in Cellular Neuroscience*, *13*(528). https://doi.org/10.3389/fncel.2019.00528

Northstone, K., Lewcock, M., Groom, A., Boyd, A., Macleod, J., Timpson, N., & Wells, N. (2019). The Avon Longitudinal Study of Parents and Children (ALSPAC): an update on the enrolled sample of index children in 2019. *Wellcome open research*, *4*.

Novick, M. R. (1966). The axioms and principal results of classical test theory. *Journal of mathematical psychology*, *3*(1), 1-18.

Osborn, A. J., Mathias, J. L., & Fairweather-Schmidt, A. K. (2014). Depression following adult, non-penetrating traumatic brain injury: A meta-analysis examining methodological variables and sample characteristics. *Neuroscience and Biobehavioral Reviews*, *47*, 1-15. https://doi.org/10.1016/j.neubiorev.2014.07.007

Osborn, A. J., Mathias, J. L., & Fairweather-Schmidt, A. K. (2016). Prevalence of Anxiety Following Adult Traumatic Brain Injury: A Meta-Analysis Comparing Measures, Samples and Postinjury Intervals. *Neuropsychology*, *30*(2), 247-261. https://doi.org/10.1037/neu0000221

Osborne-Crowley, K., Wilson, E., De Blasio, F., Wearne, T., Rushby, J., & McDonald, S. (2019). Preserved rapid conceptual processing of emotional expressions despite reduced neuropsychological performance following traumatic brain injury. *Neuropsychology*. https://doi.org/10.1037/neu0000545

Paiva-Silva, A. I. d., Pontes, M. K., Aguiar, J. S. R., & de Souza, W. C. (2016). How do we evaluate facial emotion recognition? *Psychology & neuroscience*, *9*(2), 153.

Palm, M. E., Elliott, R., McKie, S., Deakin, J. F. W., & Anderson, I. M. (2011). Attenuated responses to emotional expressions in women with generalized anxiety disorder. *Psychological Medicine*, *41*(5), 1009-1018. https://doi.org/10.1017/s0033291710001455

Palmer, C. E., Langbehn, D., Tabrizi, S. J., & Papoutsi, M. (2018). Test Retest Reliability of Measures Commonly Used to Measure Striatal Dysfunction across Multiple Testing Sessions: A Longitudinal Study. *Frontiers in Psychology*, *8*, Article 2363. https://doi.org/10.3389/fpsyg.2017.02363

Parker, J. D. A., Taylor, G. J., & Bagby, R. M. (1993). ALEXITHYMIA AND THE RECOGNITION OF FACIAL EXPRESSIONS OF EMOTION [Article]. *Psychotherapy and Psychosomatics*, *59*(3-4), 197-202. https://doi.org/10.1159/000288664

Parker, J. D. A., Taylor, G. J., & Bagby, R. M. (2003). The 20-item Toronto Alexithymia Scale - III. Reliability and factorial validity in a community population [Article]. *Journal of psychosomatic research*, *55*(3), 269-275. https://doi.org/10.1016/s0022-3999(02)00578-0

Parkinson, B. (2005). Do facial movements express emotions or communicate motives? *Personality and Social Psychology Review*, *9*(4), 278-311. https://doi.org/10.1207/s15327957pspr0904_1

Parsons, S. (2020). *Calculate Task Split Half Reliability Estimates*. In (Version 0.7.1) https://github.com/sdparsons/splithalf

Parsons, S., Kruijt, A.-W., & Fox, E. (2019). Psychological Science needs a standard practice of reporting the reliability of cognitive behavioural measurements. *Advances in Methods and Practices in Psychological Science*.

Patton, G. C., Coffey, C., Posterino, M., Carlin, J. B., Wolfe, R., & Bowes, G. (1999). A computerised screening instrument for adolescent depression: population-based validation and application to a two-phase case-control study. *Social psychiatry and psychiatric epidemiology*, *34*(3), 166-172. https://doi.org/10.1007/s001270050129

Peer, E., Brandimarte, L., Samat, S., & Acquisti, A. (2017). Beyond the Turk: Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology*, *70*, 153-163. https://doi.org/10.1016/j.jesp.2017.01.006

Peeters, W., Majdan, M., Brazinova, A., Nieboer, D., & Maas, A. I. R. (2017). Changing Epidemiological Patterns in Traumatic Brain Injury: A Longitudinal Hospital-Based Study in Belgium. *Neuroepidemiology*, *48*(1-2), 63-70. https://doi.org/10.1159/000471877

Penton-Voak, I. S., Bate, H., Lewis, G., & Munafo, M. R. (2012). Effects of emotion perception training on mood in undergraduate students: randomised controlled trial. *Br J Psychiatry*, *201*(1), 71-72. https://doi.org/10.1192/bjp.bp.111.107086

Penton-Voak, I. S., Munafò, M. R., & Looi, C. Y. (2017). Biased Facial-Emotion Perception in Mental Health Disorders: A Possible Target for Psychological Intervention? *Current Directions in Psychological Science*, *26*(3), 294-301. https://doi.org/10.1177/0963721417704405

Penton-Voak, I. S., Thomas, J., Gage, S. H., McMurran, M., McDonald, S., & Munafo, M. R. (2013). Increasing recognition of happiness in ambiguous facial expressions reduces anger and aggressive behavior. *Psychol Sci*, *24*(5), 688-697. https://doi.org/10.1177/0956797612459657

Phyland, R. K., Ponsford, J., Carrier, S. L., Hicks, A. J., & McKay, A. (2021). Agitated Behaviors following Traumatic Brain Injury: A Systematic Review and Meta-Analysis of Prevalence by Post-Traumatic Amnesia Status, Hospital Setting, and Agitated Behavior Type. *Journal of Neurotrauma*, *38*(22), 3047-3067. https://doi.org/10.1089/neu.2021.0257

Plana, I., Lavoie, M. A., Battaglia, M., & Achim, A. M. (2014). A meta-analysis and scoping review of social cognition performance in social phobia, posttraumatic stress disorder and other anxiety disorders [Review]. *Journal of Anxiety Disorders*, *28*(2), 169-177. https://doi.org/10.1016/j.janxdis.2013.09.005

Polit, D. F. (2014). Getting serious about test–retest reliability: a critique of retest research and some recommendations. *Quality of Life Research*, *23*(6), 1713-1720. https://link.springer.com/content/pdf/10.1007%2Fs11136-014-0632-9.pdf

Ponsford, J., Alway, Y., & Gould, K. R. (2018). Epidemiology and Natural History of Psychiatric Disorders After TBI. *Journal of Neuropsychiatry and Clinical*

Neurosciences, *30*(4), 262-270. https://doi.org/10.1176/appi.neuropsych.18040093

Ponsford, J., Downing, M. G., Olver, J., Ponsford, M., Acher, R., Carty, M., & Spitz, G. (2014). Longitudinal Follow-Up of Patients with Traumatic Brain Injury: Outcome at Two, Five, and Ten Years Post-Injury. *Journal of Neurotrauma*, *31*(1), 64-77. https://doi.org/10.1089/neu.2013.2997

Ponterotto, J. G., & Ruckdeschel, D. E. (2007). An overview of coefficient alpha and a reliability matrix for estimating adequacy of internal consistency coefficients with psychological research measures. *Perceptual and Motor Skills*, *105*(3), 997-1014. https://doi.org/10.2466/pms.105.3.997-1014

Prince, C., & Bruhns, M. E. (2017). Evaluation and Treatment of Mild Traumatic Brain Injury: The Role of Neuropsychology. *Brain Sciences*, *7*(8), Article 105. https://doi.org/10.3390/brainsci7080105

Prkachin, G. C., Casey, C., & Prkachin, K. M. (2009). Alexithymia and perception of facial expressions of emotion. *Personality and Individual Differences*, *46*(4), 412-417. https://doi.org/https://doi.org/10.1016/j.paid.2008.11.010

Prochnow, D., Donell, J., Schafer, R., Jorgens, S., Hartung, H. P., Franz, M., & Seitz, R. J. (2011). Alexithymia and impaired facial affect recognition in multiple sclerosis. *Journal of Neurology*, *258*(9), 1683-1688. https://doi.org/10.1007/s00415-011-6002-4

R Core Team. (2020). *R: A language and environment for statistical computing*. In (Version 4.0.2) R Foundation for Statistical Computing. http://www.R-project.org/

Ramos, S. D., Liddement, J., Addicott, C., Fortescue, D., & Oddy, M. (2020). The development of the Brain Injury Screening Index (BISI): A self-report measure [Article]. *Neuropsychological Rehabilitation*, *30*(5), 948-960. https://doi.org/10.1080/09602011.2018.1526692

Ramsay, M. C., & Reynolds, C. R. (1995). Separate digits tests: A brief history, a literature review, and a reexamination of the factor structure of the test of memory and learning (TOMAL) [Article]. *Neuropsychology Review*, *5*(3), 151-171. https://doi.org/10.1007/bf02214760

Rao, V., Rosenberg, P., Bertrand, M., Salehinia, S., Spiro, J., Vaishnavi, S., Rastogi, P., Noll, K., Schretlen, D. J., Brandt, J., Cornwell, E., Makley, M., & Miles, Q. S. (2009). Aggression After Traumatic Brain Injury: Prevalence and Correlates. *Journal of Neuropsychiatry and Clinical Neurosciences*, *21*(4), 420-429. <Go to ISI>://WOS:000272311300009

Reis, H. T., & Judd, C. M. (2000). *Handbook of research methods in social and personality psychology*. Cambridge University Press.

Richards, A., French, C. C., Calder, A. J., Webb, B., Fox, R., & Young, A. W. (2002). Anxiety-Related Bias in the Classification of Emotionally Ambiguous Facial Expressions [Article]. *Emotion*, *2*(3), 273-287. https://doi.org/10.1037/1528-3542.2.3.273

Rigon, A., Turkstra, L., Mutlu, B., & Duff, M. (2016). The female advantage: sex as a possible protective factor against emotion recognition impairment following traumatic brain injury. *Cognitive Affective & Behavioral Neuroscience*, *16*(5), 866-875. https://doi.org/10.3758/s13415-016-0437-0

Rigon, A., Voss, M. W., Turkstra, L. S., Mutlu, B., & Duff, M. C. (2017). Relationship between individual differences in functional connectivity and facial-emotion recognition abilities in adults with traumatic brain injury. *Neuroimage-Clinical*, *13*, 370-377. https://doi.org/10.1016/j.nicl.2016.12.010

Rizopoulos, D. (2006). ltm: An R package for Latent Variable Modelling and Item Response Theory Analyses. *Journal of Statistical Software*, *17*, 1--25. http://www.jstatsoft.org/v17/i05/

Rosen, H. J., & Levenson, R. W. (2009). The emotional brain: combining insights from patients and basic science. *Neurocase*, *15*(3), 173-181. https://doi.org/10.1080/13554790902796787

Rosenberg, H., Dethier, M., Kessels, R. P. C., Westbrook, R. F., & McDonald, S. (2015). Emotion perception after moderate-severe traumatic brain injury: The valence effect and the role of working memory, processing speed, and nonverbal reasoning. *Neuropsychology*, *29*(4), 509-521. https://doi.org/10.1037/neu0000171

Rosenberg, H., McDonald, S., Dethier, M., Kessels, R. P. C., & Westbrook, R. F. (2014). Facial Emotion Recognition Deficits following Moderate-Severe Traumatic Brain Injury (TBI): Re-examining the Valence Effect and the Role of Emotion Intensity [Article]. *Journal of the International Neuropsychological Society*, *20*(10), 994-1003. https://doi.org/10.1017/s1355617714000940

Rosenberg, H., McDonald, S., Rosenberg, J., & Westbrook, R. F. (2018). Amused, flirting or simply baffled? Is recognition of all emotions affected by traumatic brain injury? [Article]. *Journal of Neuropsychology*, *12*(2), 145-164. https://doi.org/10.1111/jnp.12109

Rosenberg, H., McDonald, S., Rosenberg, J., & Westbrook, R. F. (2019). Measuring emotion perception following traumatic brain injury: The Complex Audio Visual Emotion Assessment Task (CAVEAT) [Article]. *Neuropsychological Rehabilitation*, *29*(2), 232-250. https://doi.org/10.1080/09602011.2016.1273118

Rosenberg, N., Ihme, K., Lichev, V., Sacher, J., Rufer, M., Grabe, H. J., Kugel, H., Pampel, A., Lepsien, J., Kersting, A., Villringer, A., & Suslow, T. (2020). Alexithymia and automatic processing of facial emotions: behavioral and neural findings [Article]. *Bmc Neuroscience*, *21*(1), 13, Article 23. https://doi.org/10.1186/s12868-020-00572-6

Rossi, R., Zammit, S., Button, K. S., Munafo, M. R., Lewis, G., & David, A. S. (2016). Psychotic Experiences and Working Memory: A Population-Based Study Using Signal-Detection Analysis. *PLoS One*, *11*(4), Article e0153148. https://doi.org/10.1371/journal.pone.0153148

Rossignol, M., Philippot, P., Douilliez, C., Crommelinck, M., & Campanella, S. (2005). The perception of fearful and happy facial expression is modulated by anxiety: an event-related potential study. *Neuroscience Letters*, *377*(2), 115-120. https://doi.org/10.1016/j.neulet.2004.11.091

Roy, D., Vaishnavi, S., Han, D. F., & Rao, V. (2017). Correlates and Prevalence of Aggression at Six Months and One Year After First-Time Traumatic Brain Injury. *Journal of Neuropsychiatry and Clinical Neurosciences*, *29*(4), 334-342. https://doi.org/10.1176/appi.neuropsych.16050088

RStudio Team. (2020). *RStudio: Integrated Development Environment for R*. In RStudio, PBC. http://www.rstudio.com/

Ruff, R. M., Iverson, G. L., Barth, J. T., Bush, S. S., Broshek, D. K., Policy, N. A. N., & Planning, C. (2009). Recommendations for Diagnosing a Mild Traumatic Brain Injury: A National Academy of Neuropsychology Education Paper [Editorial Material]. *Archives of Clinical Neuropsychology*, *24*(1), 3-10. https://doi.org/10.1093/arclin/acp006

Ruffman, T., Henry, J. D., Livingstone, V., & Phillips, L. H. (2008). A meta-analytic review of emotion recognition and aging: implications for neuropsychological models of

aging. *Neurosci Biobehav Rev*, *32*(4), 863-881. https://doi.org/10.1016/j.neubiorev.2008.01.001

Russo, M., Mahon, K., Shanahan, M., Solon, C., Ramjas, E., Turpin, J., & Burdick, K. E. (2015). The association between childhood trauma and facial emotion recognition in adults with bipolar disorder. *Psychiatry research*, *229*(3), 771-776. https://doi.org/10.1016/j.psychres.2015.08.004

Sabaz, M., Simpson, G. K., Walker, A. J., Rogers, J. M., Gillis, I., & Strettles, B. (2014). Prevalence, Comorbidities, and Correlates of Challenging Behavior Among Community-Dwelling Adults With Severe Traumatic Brain Injury: A Multicenter Study. *Journal of Head Trauma Rehabilitation*, *29*(2), E19-E30. https://doi.org/10.1097/HTR.0b013e31828dc590

Sasson, N. J., Pinkham, A. E., Richard, J., Hughett, P., Gur, R. E., & Gur, R. C. (2010). Controlling for Response Biases Clarifies Sex and Age Differences in Facial Affect Recognition. *Journal of Nonverbal Behavior*, *34*(4), 207-221. https://doi.org/10.1007/s10919-010-0092-z

Sattler, J. M. (2001). *Assessment of children: Cognitive applications*. Jerome M Sattler Publisher.

Schmidt, A. F., & Finan, C. (2018). Linear regression and the normality assumption. *Journal of Clinical Epidemiology*, *98*, 146-151. https://doi.org/https://doi.org/10.1016/j.jclinepi.2017.12.006

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-prime computer software and manual*. In Psychology Software Tools Inc.

Schweitzer, A. D., Niogi, S. N., Whitlow, C. J., & Tsiouris, A. J. (2019). Traumatic Brain Injury: Imaging Patterns and Complications [Article]. *Radiographics*, *39*(6), 1571-+. https://doi.org/10.1148/rg.2019190076

Schyns, P. G., Petro, L. S., & Smith, M. L. (2009). Transmission of Facial Expressions of Emotion Co-Evolved with Their Efficient Decoding in the Brain: Behavioral and Brain Evidence. *PLoS One*, *4*(5), Article e5625. https://doi.org/10.1371/journal.pone.0005625

Sherman, E., Brooks, B. L., Iverson, G. L., Slick, D. J., & Strauss, E. (2011). Reliability and validity in neuropsychology. In *The little black book of neuropsychology* (pp. 873-892). Springer.

Shrout, P. E., & Fleiss, J. L. (1979). INTRACLASS CORRELATIONS - USES IN ASSESSING RATER RELIABILITY. *Psychological Bulletin*, *86*(2), 420-428. https://doi.org/10.1037/0033-2909.86.2.420

Si, B., Dumkrieger, G., Wu, T., Zafonte, R., Dodick, D. W., Schwedt, T. J., & Li, J. (2018). A Cross-Study Analysis for Reproducible Sub-classification of Traumatic Brain Injury [Article]. *Frontiers in Neurology*, *9*, 12, Article 606. https://doi.org/10.3389/fneur.2018.00606

Smith, M. L. (2012). Rapid Processing of Emotional Expressions without Conscious Awareness. *Cerebral Cortex*, *22*(8), 1748-1760. https://doi.org/10.1093/cercor/bhr250

Soveri, A., Lehtonen, M., Karlsson, L. C., Lukasik, K., Antfolk, J., & Laine, M. (2018). Test-retest reliability of five frequently used executive tasks in healthy adults. *Applied Neuropsychology-Adult*, *25*(2), 155-165. https://doi.org/10.1080/23279095.2016.1263795

Spearman, C. (1904). The Proof and Measurement of Association between Two Things. *The American Journal of Psychology*, *15*(1), 72-101.

Spielberger, C. D. (1983). State-trait anxiety inventory for adults.

Spikman, J. M., Boelen, D. H., Pijnenborg, G. H., Timmerman, M. E., van der Naalt, J., & Fasotti, L. (2013). Who benefits from treatment for executive dysfunction after

brain injury? Negative effects of emotion recognition deficits. *Neuropsychol Rehabil*, *23*(6), 824-845. https://doi.org/10.1080/09602011.2013.826138

Spikman, J. M., Milders, M. V., Visser-Keizer, A. C., Westerhof-Evers, H. J., Herben-Dekker, M., & van der Naalt, J. (2013). Deficits in Facial Emotion Recognition Indicate Behavioral Changes and Impaired Self-Awareness after Moderate to Severe Traumatic Brain Injury. *PLoS One*, *8*(6), Article e65581. https://doi.org/10.1371/journal.pone.0065581

Spikman, J. M., Timmerman, M. E., Milders, M. V., Veenstra, W. S., & van der Naalt, J. (2012). Social Cognition Impairments in Relation to General Cognitive Deficits, Injury Severity, and Prefrontal Lesions in Traumatic Brain Injury Patients. *Journal of Neurotrauma*, *29*(1), 101-111. https://doi.org/10.1089/neu.2011.2084

St Clair-Thompson, H. L. (2010). Backwards digit recall: A measure of short-term memory or working memory? [Article]. *European Journal of Cognitive Psychology*, *22*(2), 286-296. https://doi.org/10.1080/09541440902771299

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures [Article]. *Behavior Research Methods Instruments & Computers*, *31*(1), 137-149. https://doi.org/10.3758/bf03207704

Stanley, D. (2018). *Create American Psychological Association (APA) Style Tables*. In (Version 2.0.5) https://github.com/dstanley4/apaTables

Steiger, J. H. (1980). TESTS FOR COMPARING ELEMENTS OF A CORRELATION MATRIX [Article]. *Psychological Bulletin*, *87*(2), 245-251. https://doi.org/10.1037/0033-2909.87.2.245

Sterne, J. A. C., & Smith, G. D. (2001). Sifting the evidence - what's wrong with significance tests? *Bmj-British Medical Journal*, *322*(7280), 226-+. https://doi.org/10.1136/bmj.322.7280.226

Steyerberg, E. W., Wiegers, E., Sewalt, C., Buki, A., Citerio, G., De Keyser, V., Ercole, A., Kunzmann, K., Lanyon, L., Lecky, F., Lingsma, H., Manley, G., Nelson, D., Peul, W., Stocchetti, N., von Steinbuchel, N., Vande Vyvere, T., Verheyden, J., Wilson, L., Maas, A. I. R., Menon, D. K., & Investigat, C.-T. P. (2019). Case-mix, care pathways, and outcomes in patients with traumatic brain injury in CENTER-TBI: a European prospective, multicentre, longitudinal, cohort study [Article]. *Lancet Neurology*, *18*(10), 923-934. https://doi.org/10.1016/s1474-4422(19)30232-7

Stone, C. (2019). A Defense and Definition of Construct Validity in Psychology [Article]. *Philosophy of Science*, *86*(5), 1250-1261. https://doi.org/10.1086/705567

Strauss, E., Sherman, E. M., & Spreen, O. (2006). *A compendium of neuropsychological tests: Administration, norms, and commentary*. American Chemical Society.

Streiner, D. L., Norman, G. R., & Cairney, J. (2008). *Health measurement scales: a practical guide to their development and use* (Fourth Edition ed.). Oxford University Press, USA. https://doi.org/10.1093/acprof:oso/9780199231881.001.0001

Streiner, D. L., Norman, G. R., & Cairney, J. (2015). *Health Measurement Scales : A Practical Guide to Their Development and Use*. Oxford University Press. http://ebookcentral.proquest.com/lib/bristol/detail.action?docID=1816173

Suer, M., & Abd-Elsayed, A. (2020). Patient with Traumatic Brain Injury. In A. Abd-Elsayed (Ed.), *Guide to the Inpatient Pain Consult* (pp. 429-443). Springer International Publishing. https://doi.org/10.1007/978-3-030-40449-9_29

Surcinelli, P., Codispoti, M., Montebarocci, O., Rossi, N., & Baldaro, B. (2006). Facial emotion recognition in trait anxiety. *J Anxiety Disord*, *20*(1), 110-117. https://doi.org/10.1016/j.janxdis.2004.11.010

Suslow, T., Husslack, A., Bujanow, A., Henkelmann, J., Kersting, A., Hoffmann, K. T., Egloff, B., Lobsien, D., & Gunther, V. (2019). Implicitly and explicitly assessed

anxiety: No relationships with recognition of and brain response to facial emotions [Article]. *Neuroscience*, *408*, 1-13. https://doi.org/10.1016/j.neuroscience.2019.03.059

Suzuki, A., Hoshino, T., & Shigemasu, K. (2006). Measuring individual differences in sensitivities to basic emotions in faces. *Cognition*, *99*(3), 327-353. https://doi.org/10.1016/j.cognition.2005.04.003

Svingos, A. M., Asken, B. M., Jaffee, M. S., Bauer, R. M., & Heaton, S. C. (2019). Predicting long-term cognitive and neuropathological consequences of moderate to severe traumatic brain injury: Review and theoretical framework. *Journal of Clinical and Experimental Neuropsychology*, *41*(8), 775-785. https://doi.org/10.1080/13803395.2019.1620695

Sweeny, T. D., Suzuki, S., Grabowecky, M., & Paller, K. A. (2013). Detecting and Categorizing Fleeting Emotions in Faces. *Emotion*, *13*(1), 76-91. https://doi.org/10.1037/a0029193

Tateno, A., Jorge, R. E., & Robinson, R. G. (2003). Clinical correlates of aggressive behavior after traumatic brain injury. *Journal of Neuropsychiatry and Clinical Neurosciences*, *15*(2), 155-160. https://doi.org/10.1176/appi.neuropsych.15.2.155

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International journal of medical education*, *2*, 53-55. https://doi.org/10.5116/ijme.4dfb.8dfd

Taylor, G. J., & Bagby, R. M. (2004). New trends in alexithymia research [Review]. *Psychotherapy and Psychosomatics*, *73*(2), 68-77. https://doi.org/10.1159/000075537

Teasdale, G., & Jennett, B. (1974). ASSESSMENT OF COMA AND IMPAIRED CONSCIOUSNESS - PRACTICAL SCALE. *Lancet*, *2*(7872), 81-84. <Go to ISI>://WOS:A1974T535500009

Teasdale, G., & Jennett, B. (1976). ASSESSMENT AND PROGNOSIS OF COMA AFTER HEAD-INJURY. *Acta Neurochirurgica*, *34*(1-4), 45-55. https://doi.org/10.1007/bf01405862

Theadom, A., McDonald, S., Starkey, N., Barker-Collo, S., Jones, K. M., Ameratunga, S., Wilson, E., Feigin, V. L., & Grp, B. I. y. R. (2019). Social Cognition Four Years After Mild-TBI: An Age-Matched Prospective Longitudinal Cohort Study. *Neuropsychology*, *33*(4), 560-567. https://doi.org/10.1037/neu0000516

Thompson, A. E., & Voyer, D. (2014). Sex differences in the ability to recognise non-verbal displays of emotion: A meta-analysis. *Cognition & Emotion*, *28*(7), 1164-1195. https://doi.org/10.1080/02699931.2013.875889

Turkstra, L. S., Mutlu, B., Ryan, C. W., Stafslien, E. D. H., Richmond, E. K., Hosokawa, E., & Duff, M. C. (2020). Sex and Gender Differences in Emotion Recognition and Theory of Mind After TBI: A Narrative Review and Directions for Future Research. *Frontiers in Neurology*, *11*, Article 59. https://doi.org/10.3389/fneur.2020.00059

Uljarevic, M., & Hamilton, A. (2013). Recognition of Emotions in Autism: A Formal Meta-Analysis. *Journal of Autism and Developmental Disorders*, *43*(7), 1517-1526. https://doi.org/10.1007/s10803-012-1695-5

Vacha-Haase, T., Henson, R. K., & Caruso, J. C. (2002). Reliability generalization: Moving toward improved understanding and use of score reliability. *Educational and psychological measurement*, *62*(4), 562-569.

VanderWeele, T. J., & Mathur, M. B. (2019). SOME DESIRABLE PROPERTIES OF THE BONFERRONI CORRECTION: IS THE BONFERRONI CORRECTION REALLY SO BAD? *American Journal of Epidemiology*, *188*(3), 617-618. https://doi.org/10.1093/aje/kwy250

Vaz, S., Falkmer, T., Passmore, A. E., Parsons, R., & Andreou, P. (2013). The case for using the repeatability coefficient when calculating test-retest reliability. *PLoS One*, *8*(9), e73990. https://doi.org/10.1371/journal.pone.0073990

Venkatesan, U. M., Lancaster, K., Lengenfelder, J., & Genova, H. M. (2021). Independent contributions of social cognition and depression to functional status after moderate or severe traumatic brain injury. *Neuropsychological Rehabilitation*, *31*(6), 954-970. https://doi.org/10.1080/09602011.2020.1749675

Verbruggen, F., & Logan, G. D. (2008). Response inhibition in the stop-signal paradigm. *Trends in Cognitive Sciences*, *12*(11), 418-424. https://doi.org/10.1016/j.tics.2008.07.005

Vetter, T., & Walker, M. (2011). Computer-generated images in face perception. *Oxford handbook of face perception*, 388-399.

Vitoratou, S., & Pickles, A. (2017). A note on contemporary psychometrics. In: Taylor & Francis.

Vuilleumier, P., & Righart, R. (2011). Attention and automaticity in processing facial expressions. *Oxford handbook of face perception*, 449-478.

Wagner, H. L. (1993). ON MEASURING PERFORMANCE IN CATEGORY JUDGMENT STUDIES OF NONVERBAL BEHAVIOR [Review]. *Journal of Nonverbal Behavior*, *17*(1), 3-28. https://doi.org/10.1007/bf00987006

Wallace, E. J., Mathias, J. L., & Ward, L. (2018). Diffusion tensor imaging changes following mild, moderate and severe adult traumatic brain injury: a meta-analysis. *Brain Imaging and Behavior*, *12*(6), 1607-1621. https://doi.org/10.1007/s11682-018-9823-2

Wallis, K., Kelly, M., McRae, S. E., McDonald, S., & Campbell, L. E. (2021). Domains and measures of social cognition in acquired brain injury: A scoping review. *Neuropsychological Rehabilitation*, 1-35. https://doi.org/10.1080/09602011.2021.1933087

Walter, S. L., Seibert, S. E., Goering, D., & O'Boyle, E. H. (2019). A Tale of Two Sample Sources: Do Results from Online Panel Data and Conventional Data Converge? *Journal of Business and Psychology*, *34*(4), 425-452. https://doi.org/10.1007/s10869-018-9552-y

Watters, A. J., & Williams, L. M. (2011). NEGATIVE BIASES AND RISK FOR DEPRESSION; INTEGRATING SELF-REPORT AND EMOTION TASK MARKERS. *Depression and Anxiety*, *28*(8), 703-718. https://doi.org/10.1002/da.20854

Weafer, J., Baggott, M. J., & de Wit, H. (2013). Test-Retest Reliability of Behavioral Measures of Impulsive Choice, Impulsive Action, and Inattention. *Experimental and Clinical Psychopharmacology*, *21*(6), 475-481. https://doi.org/10.1037/a0033659

Wearne, T., Osborne-Crowley, K., Rosenberg, H., Dethier, M., & McDonald, S. (2019). Emotion recognition depends on subjective emotional experience and not on facial expressivity: evidence from traumatic brain injury [Article]. *Brain Injury*, *33*(1), 12-22. https://doi.org/10.1080/02699052.2018.1531300

Webb, N. M., Shavelson, R. J., & Haertel, E. H. (2006). 4 reliability coefficients and generalizability theory. *Handbook of statistics*, *26*, 81-124.

Webster, M. A., Kaping, D., Mizokami, Y., & Duhamel, P. (2004). Adaptation to natural facial categories. *Nature*, *428*(6982), 557-561. https://doi.org/10.1038/nature02420

Weir, J. (2005). Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *The Journal of Strength Conditioning Research*, *19*(1), 231-240.

West, J. T., Horning, S. M., Klebe, K. J., Foster, S. M., Cornwell, R. E., Perrett, D., Burt, D. M., & Davis, H. P. (2012). Age effects on emotion recognition in facial displays: from 20 to 89 years of age. *Exp Aging Res*, *38*(2), 146-168. https://doi.org/10.1080/0361073X.2012.659997

Westerhof-Evers, H. J., Visser-Keizer, A. C., McDonald, S., & Spikman, J. M. (2014). Performance of healthy subjects on an ecologically valid test for social cognition: The short, Dutch Version of The Awareness of Social Inference Test (TASIT) [Article]. *Journal of Clinical and Experimental Neuropsychology*, *36*(10), 1031-1041. https://doi.org/10.1080/13803395.2014.966661

White, N., Forsyth, B., Lee, A., & Machado, L. (2018). Repeated Computerized Cognitive Testing: Performance Shifts and Test-Retest Reliability in Healthy Young Adults [Article]. *Psychological Assessment*, *30*(4), 539-549. https://doi.org/10.1037/pas0000503

Williams, C., & Wood, R. L. (2010). Alexithymia and emotional empathy following traumatic brain injury [Article]. *Journal of Clinical and Experimental Neuropsychology*, *32*(3), 259-267. https://doi.org/10.1080/13803390902976940

Williams, C., Wood, R. L., & Howe, H. (2018). Alexithymia is associated with aggressive tendencies following traumatic brain injury. *Brain Inj*, 1-9. https://doi.org/10.1080/02699052.2018.1531302

Williams, E., Thomas, K., Sidebotham, H., & Emond, A. (2008). Prevalence and characteristics of autistic spectrum disorders in the ALSPAC cohort. *Developmental medicine & child neurology*, *50*(9), 672-677.

Worthington, A., & Wood, R. L. (2018). Apathy following traumatic brain injury: A review [Review]. *Neuropsychologia*, *118*, 40-47. https://doi.org/10.1016/j.neuropsychologia.2018.04.012

Wostmann, N. M., Aichert, D. S., Costa, A., Rubia, K., Moller, H. J., & Ettinger, U. (2013). Reliability and plasticity of response inhibition and interference control. *Brain and Cognition*, *81*(1), 82-94. https://doi.org/10.1016/j.bandc.2012.09.010

Wu, M., Kujawa, A., Lu, L. H., Fitzgerald, D. A., Klumpp, H., Fitzgerald, K. D., Monk, C. S., & Phan, K. L. (2016). Age-related changes in amygdala-frontal connectivity during emotional face processing from childhood into young adulthood. *Hum Brain Mapp*, *37*(5), 1684-1695. https://doi.org/10.1002/hbm.23129

Xiao, H., Jacobsen, A., Chen, Z. Q., & Wang, Y. (2017). Detecting social-cognitive deficits after traumatic brain injury: An ALE meta-analysis of fMRI studies. *Brain Injury*, *31*(10), 1331-1339. https://doi.org/10.1080/02699052.2017.1319576

Yassin, W., Callahan, B. L., Ubukata, S., Sugihara, G., Murai, T., & Ueda, K. (2017). Facial emotion recognition in patients with focal and diffuse axonal injury. *Brain Injury*, *31*(5), 624-630. https://doi.org/10.1080/02699052.2017.1285052

Yeates, K. O., Levin, H. S., & Ponsford, J. (2017). The Neuropsychology of Traumatic Brain Injury: Looking Back, Peering Ahead. *Journal of the International Neuropsychological Society*, *23*(9-10), 806-817. https://doi.org/10.1017/s1355617717000686

Yim, J., Babbage, D. R., Zupan, B., Neumann, D., & Willer, B. (2013). The relationship between facial affect recognition and cognitive functioning after traumatic brain injury. *Brain Injury*, *27*(10), 1155-1161. https://doi.org/10.3109/02699052.2013.804203

Yoon, K. L., & Zinbarg, R. E. (2008). Interpreting neutral faces as threatening is a default mode for socially anxious individuals. *J Abnorm Psychol*, *117*(3), 680-685. https://doi.org/10.1037/0021-843X.117.3.680

Young, A., Perrett, D., Calder, A., Sprengelmeyer, R., & Ekman, P. (2002). Facial expressions of emotion: Stimuli and tests (FEEST). *Bury St. Edmunds: Thames Valley Test Company*.

Yuvaraj, R., Murugappan, M., Norlinah, M. I., Sundaraj, K., & Khairiyah, M. (2013). Review of Emotion Recognition in Stroke Patients. *Dementia and Geriatric Cognitive Disorders*, *36*(3-4), 179-196. https://doi.org/10.1159/000353440

Zimmerman, D. W., & Zumbo, B. D. (2015). Resolving the issue of how reliability is related to statistical power: adhering to mathematical definitions. *Journal of Modern Applied Statistical Methods*, *14*(2), 5.

Zupan, B., Babbage, D., Neumann, D., & Willer, B. (2014). Recognition of facial and vocal affect following traumatic brain injury. *Brain Injury*, *28*(8), 1087-1095. https://doi.org/10.3109/02699052.2014.901560

Zupan, B., Babbage, D., Neumann, D., & Willer, B. (2017). Sex Differences in Emotion Recognition and Emotional Inferencing Following Severe Traumatic Brain Injury. *Brain Impairment*, *18*(1), 36-48. https://doi.org/10.1017/BrImp.2016.22

Chapter 8 Appendices

8.1 Appendix A (Chapter 2) - Reliability of the N-back and Stop Signal Task

8.1.1.1 N-back

The N-back task is a widely used measure of working memory and many variations of the task exist. In this study a visuoverbal 2-back version of the task was used. This means that participants were presented with visual stimuli (numbers from 0 to 9) and for each trial and were asked to indicate whether the current stimulus was the same or different from the stimulus presented two trials ago (For details please refer to section 2.2.2.2). Both reaction time and accuracy can be used as outcome measures, but this study will focus on the test-retest reliability of accuracy.

N-back tasks have been used extensively in research and clinical settings since their introduction by Kirchner in 1958. However, there is comparatively little information regarding the psychometric properties of N-back tasks (Jaeggi et al., 2010), and most studies investigating the reliability of the N-back tasks have focused on internal consistency as opposed to test-retest reliability (Jaeggi et al., 2010; Soveri et al., 2018). Hockey and Geffen (2004) reported test-retest reliability of accuracy on a visuospatial version of the 2-back task as r = 0.538. Their data also showed a trend towards increased reliability of accuracy as cognitive load increased, i.e. 3-back task has the highest reliability coefficient in their study. They suggest this may be due to ceiling effects leading to homogeneity during the lower cognitive load tasks. White and colleagues (2018) assessed performance on a 2-back visuoverbal task at 6 timepoints and calculated test-retest reliability coefficients (ICC2.1) between these timepoints. They reported moderate to good reliability with coefficients ranging from 0.63 to 0.87 between the various timepoints. A limitation of their study is that they did not report the confidence intervals around these estimates, which have previously been shown to be very wide for the N-back tasks (Soveri et al., 2018).

Reliability estimates for both agreement in scores between testing sessions and consistency of performance in relation to other participants were low on the visuoverbal 2-back version of the N-back task. The reliability coefficients calculated for the 2-back task in this study are slightly lower than those reported in the current literature (Hockey & Geffen, 2004; White et al., 2018). The 3-back version of the task was associated with higher reliability estimates than the 2-back version in a study by Hockey and Geffen

(2004). This could indicate a ceiling effect on the 2-back version of the task, which would result in homogeneity of performance and thus decrease reliability estimates. This might also explain why we did not detect a difference in performance between test session 1 and session 2. The SEM and SRD, however, were relatively large suggesting that there was variation in individual scores, despite there being no evidence for a difference in performance between testing sessions. Further, results from a recent study by Soveri and colleagues (2018) showed that despite the reliability coefficient for a visuoverbal 3-back task indicating moderate reliability (ICC2.1 = 0.567) the confidence intervals around their estimate were large (95% CI = 0.074 – 0.819). Prior studies may have been overestimating test-retest reliability for the 2-back task as confidence intervals around the reliability estimates were not reported or considered. The current findings are indicative that the N-back task may not be suitable for use in an individual differences context, although it could still be useful in other contexts, for example, investigating group differences (Hedge et al., 2018).

8.1.1.2 Stop Signal Task

The Stop Signal Task (SST) was used as a measure of response inhibition associated with executive control (Verbruggen & Logan, 2008). In this study participants were asked to identify a letter presented on the screen as quickly as possible but instructed that if they heard a beep shortly after the letter was presented, they were not to respond to that trial (For details please refer to the method section). Performance is dependent on the interaction between the 'go' process of responding to a trial quickly and the 'stop' process of inhibiting a response, which have been associated with different neural pathways (Verbruggen & Logan, 2008). A commonly used measure of the interaction between these processes is a Stop Signal Reaction Time (SSRT), which can be calculated in a number of different ways (Band et al., 2003).

Studies investigating the test-retest reliability of the SST have generally reported low to moderate reliability. Weafer, Baggott and de Wit (2013) used SSRT as their primary outcome. They reported moderate to high reliability based on Persons r = 0.65. However, Wostmann and colleagues (2013) suggested very low reliability of the SST based on the SSRT in their study, with both Pearson's r of and ICC2.1 reported as 0.03. It is unclear exactly what method was used to calculate the SSRT in these studies and neither study reported 95% confidence intervals. As part of their investigation into why "robust cognitive tasks do not produce reliable individual differences", Hedge and

colleagues (2018) calculated the test-retest reliability of two types of SSRT, SSRT mean and SSRT integration. In both studies conducted they report low reliability for both SSRT measures, for SSRT mean the ICC2.1 = 0.47 (0.21-0.67) and 0.43 (0.19-0.62) and for SSRT integration the ICC2.1 = 0.36 (0.08-0.59) and 0.49 (0.26-0.66). In this study the SSRT median as outlined by Band and colleagues (2003) will be used as an outcome measure, because it could reduce variation due to noise leading to low reliability estimates.

Of the three cognitive tasks investigated in this study, the measurements made using the SST showed the highest test-retest reliability. The ICC2.1 was substantially lower than the two estimates indicating consistency, which is not surprising given the evidence for an improvement in performance between testing sessions. Much like for the Bristol ERT, reliability estimates based on consistency as opposed to agreement may be more appropriate when considering the impact SST reliability may be having on ALSPAC data because there is evidence for a systematic improvement. Test-retest reliability for the SST were slightly higher than reported in the literature despite the confidence intervals around the reliability estimates being quite large (Hedge et al., 2018; Weafer et al., 2013; Wostmann et al., 2013). A possible explanation for this discrepancy is the way in which the SSRT was calculated. Hedge and colleagues (2018) reported test-retest reliability estimates for the SSRT mean and SSRT integration, whilst the SSRT median was used as an outcome measure in this study (Band et al., 2003). Potentially the SSRT median is a more stable outcome to use when assessing individual difference and change in performance over time. A direct comparison of test-retest reliability of these three outcome measures in the same sample could provide further insight. Based on the test-retest reliability calculated in this ALSPAC sample, the SEM was quite large, and individuals would have to substantially reduce their SSRT to confidently say that there has been a change in performance. Again, the magnitude of change in individual scores is likely influenced by the systematic improvement in performance observed between sessions.

8.2 Appendix B (Chapter 2) - Effect size calculations

Spreadsheet published in the supplementary materials from 'Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs' by Daniel Lakens (2013b) was used to calculate the effect sizes for this study. The spreadsheet version 4.2 retrieved from the Open Science frame work (Lakens, 2013a).

The means, standard deviations, correlations, and sample size were used to calculate the effect size based on a t-test for a correlated (or dependant) sample. Data required to replicate these calculations are presented in Table 8-1.

Table 8-1

Data required to replicate effect size calculations

Task (outcome measure)	***n***	**Delay between sessions**	Time 1		Time 2		Correlation	Paired t-test	
			Mean	**SD**	**Mean**	**SD**	**r**	**t**	**p**
Bristol ERT (prop. of totalhits)	90	combined	0.693	0.074	0.722	0.074	0.60	4.13	<.001
	45	short	0.697	0.077	0.733	0.075	0.62	3.75	<.001
	45	long	0.690	0.072	0.711	0.073	0.58	2.11	0.04
N-back (*d*'prime)	87	combined	2.858	0.695	2.902	0.707	0.43	0.55	0.59
	44	short	2.97	0.677	3.10	0.552	0.21	1.16	0.25
	43	long	2.75	0.704	2.70	0.791	0.54	-0.45	0.65
SST ($SSRT_{med}$)	88	combined	255.3	49.03	232.8	37.42	0.73	-6.31	<.001
	43	short	253.2	45.4	231.5	35.6	0.73	-4.70	<.001
	45	long	257.5	53.0	234.1	39.6	0.73	-4.23	<.001

Based on recommendations by Lakens (2013b), Hedge's g_{av} is reported based on Cohen's d_{av}. Instead of using the standard deviations from both samples to calculate Cohen's d the average standard deviation is calculated, making it more appropriate for repeated measures samples. The Common Language Effect Size is recommended as an easily interpretable effect size also based on Cohen's d. Cohen's d is expressed as a percentage likelihood that a participant has a higher score in one measure compared to another. A 50% likelihood is equivalent to chance.

Hedges g_{av} and the common language (CL) calculated based on the data in Table 8-1 are reported in Table 8-2, the majority of which are also reported in the main text. All the effect size estimates are corrected for reliability based on Pearson's r.

Table 8-2

Effect sizes for the difference in scores between session 1 and session 2

Task (outcome measure)	**Delay between sessions**	**Hege's g av**	**Common language (CL) effect size**
Bristol ERT (prop. of totalhits)	combined	0.389	0.67
	short	0.466	0.71
	long	0.285	0.62
N-back (*d*'prime)	combined	0.062	0.52
	short	0.207	0.57
	long	0.066	0.53
SST ($SSRT_{med}$)	combined	0.511	0.75
	short	0.522	0.76
	long	0.492	0.74

8.3 Appendix C (Chapter 2) - Sensitivity analysis of only full cases

Analyses conducted using only participants that had complete data for all three cognitive tasks and both time points (n = 85). Table 8-3 shows the reliability estimates calculated for this data set. Results for the comparison of means at the first and second testing session for the primary outcome measure of each task using a paired t-test are presented in Table 8-4.

Table 8-3

Reliability estimates including only cases with full data

Task (outcome measure)	Intraclass Correlation Coefficient					Pearson's r		
	df	**ICC2.1**	**95% CI**	**ICC3.1**	**95% CI**	**df**	**r**	**95% CI**
Bristol ERT (prop. of totalhits)	84	0.58	0.38, 0.72	0.62	0.47, 0.73	83	0.62	0.47, 0.73
N-back (*d*'prime)	84	0.43	0.24, 0.58	0.43	0.24, 0.59	83	0.42	0.23, 0.58
SST ($SSRT_{med}$)	84	0.63	0.31, 0.79	0.70	0.58, 0.80	83	0.73	0.61, 0.81

Note: df is degrees of freedom. CI refers to 95% Confidence Interval.

Table 8-4

Summary of primary outcome and results from paired t-test

Task (outcome measure)	***n***	Time 1		Time 2		Welch's paired t-test		
		Mean	**SD**	**Mean**	**SD**	**t-value**	**p-value**	**g_{av}**
Bristol ERT (prop. of totalhits)	85	0.695	0.07	0.723	0.08	4.023	<.001**	0.375
N-back (*d*'prime)	85	2.86	0.70	2.92	0.71	0.68	0.5	0.084
SST ($SSRT_{med}$)	85	253.7	48.4	232.1	37.4	-5.97	.001**	0.495

*Note: * exceeds significance level of 0.05. ** exceeds significance level after Bonferroni correction. g_{av} is average Hedges' g corrected for reliability using Pearson's r*

8.4 Appendix A (Chapter 3) - Prolific screeners

1. **Age**
   Question: What is your date of birth?
   Setting: Minimum 18 years - Maximum 100 years

2. **Vision**
   Question: Do you have normal or corrected-to-normal vision? (i.e. You can see colour normally, and if you need glasses, you are wearing them or contact lenses)
   Response required: Yes

3. **Medication use**
   Question: Are you currently taking any medication to treat symptoms of depression, anxiety or low-mood (e.g. SSRIs)?
   Response required: No

4. **Mild Cognitive Impairment/Dementia**
   Question: Have you ever been diagnosed with mild cognitive impairment or dementia?
   Response required: No

5. **Autism Spectrum Disorder**
   Question: Have you received a formal clinical diagnosis of autistic spectrum disorder, made by a psychiatrist, psychologist, or other qualified medical specialist? This includes Asperger's syndrome, Autistic Disorder, High Functioning Autism or Pervasive Developmental Disorder
   Response required: No

6. **Mental health/illness/condition – ongoing**
   Question: Do you have – or have you had – a diagnosed, on-going mental health/illness/condition?
   Response required: No

7. **Chronic Disease**
   Question: Have you been diagnosed with any chronic diseases such as diabetes, heart disease, stroke etc.?
   Response required: No

8. **Head Injury: Knock out history**
   Question: Have you ever had an injury to the head that's caused you to be knocked out for a period of time (E.g. from a fall, blow to the head, road traffic accident)?
   Response required: No

9. **Multiple sclerosis**
   Question: Have you ever been diagnosed with multiple sclerosis (MS)?
   Response required: No

8.5 Appendix B (Chapter 3) – Table of Nationalities

Table 8-5

Nationalities of participants in the study

Nationality	Number of participants
Argentina	1
Australia	1
Belgium	1
Canada	6
China	1
Estonia	1
France	2
Germany	3
Greece	11
Haiti	1
Hungary	5
India	1
Italy	15
Jordan	1
Latvia	1
Mexico	3
Netherlands	2
Nigeria	1
Poland	22
Portugal	32
Slovenia	2
South Africa	1
Spain	9
Taiwan	2
Turkey	1
United Kingdom	29
United States	23
Vietnam	2
Total	**180**

8.6 Appendix C (Chapter 3) – Correlation of tasks with reduced Bristol ERT

Table 8-6

Mean, standard deviation, and correlations with confidence intervals for reduced Bristol ERT using only trials with above 40% intensity

Variable	Reduced Bristol ERT (total hits)
M (% accuracy)	50.88 (85%)
SD	4.84
1. Bristol ERT (total hits)	.92** [.90, .94]
2. Dynamic ERT (total hits)	.53** [.42, .63]
3. TASIT-S: EET (total hits)	.30** [.16, .43]
4. DSB (span length)	.14 [-.01, .28]
5. GFMT (total hits)	.29** [.15, .42]
6. TAS (total score)	-.09 [-.23, .06]

Note. M and *SD* are used to represent mean and standard deviation, respectively. Percentage accuracy is added in brackets. Values in square brackets indicate the 95% confidence interval for each correlation. The confidence interval is a plausible range of population correlations that could have caused the sample correlation (Cumming, 2014). * indicates $p < .05$. ** indicates $p < .01$.

8.7 Appendix A (Chapter 4) – Background Questionnaire given to participants

Background questionnaire given to participants at the beginning of their testing session. Control participants completed this questionnaire as part of their eligibility screening and were asked to complete two additional questions.

1. What is your date of birth? ____/____/_____
2. What is your first language? _____
3. What is your sex? _____
4. At what age did you leave education? _____
5. What is your current or former occupation? _____

Traumatic Brain Injury Screen

1. Have you ever had an injury to the head that caused you to be knocked out and/or dazed and confused, for a period of time? (E.g. from a fall, blow to the head, road traffic accident?)

YES	NO

   If yes, please explain and state how long since your injury:

   _____

   _____

   _____

2. How many times have you been knocked out and/or dazed and confused?

Once		Twice		Three times		Four times		More than four times	

   If more than four times, how many? _____

Control participants were additionally asked:

3. Have you ever had any other neurological condition?

   If yes, what was it: _____

4. Can you think of anything that might affect your ability to participate in the study?

YES	NO

   If yes, what is it: _____

8.8 Appendix B (Chapter 4) - Consent to contact form for recruitment after COVID-19

First part of the initial consent to contact form completed by clinicians if potential patient participants were interested in taking part in the study. Complete set of study documents available on Open Science Framework: https://osf.io/4b4zw/

Recruitment: Emotion Recognition after TBI Research Study

Please send to maren.muller-gloode@nbt.nhs.uk once completed.

Email address to be contacted on is a requirement.

1. **Is the patient suitable for the study?**

   Yes – Introduce the study (*main points provided on the next page*)

2. **Is the patient interested in participating?**

   Yes – Ask them to give consent to be contacted by the chief investigator with details about the study.

   Verbal consent to be contacted is given ☐

   Name of Patient:

   Email address to be contacted on:

   If they have any questions they can contact:

   Maren Müller-Glodde: maren.muller-glodde@nbt.nhs.uk

8.9 Appendix C (Chapter 4) – Correlations between variables included in this study

Table 8-7

Correlation matrix for tasks and questionnaires completed as part of this study and the age at which participants left education.

Variable	1	2	3	4	5	6	7	8	9	10
1. Bristol ERT (total hits)										
2. Education	-.06 [-.33, .21]									
3. GFMT (total hits)	**.46 [.22, .65]**	.25 [-.02, .49]								
4. DASS Total	-.11 [-.37, .17]	-.05 [-.32, .23]	.11 [-.17, .37]							
5. DASS Depression	.00 [-.27, .28]	-.03 [-.30, .25]	.17 [-.11, .42]	**.90 [.83, .94]**						
6. DASS Anxiety	-.26 [-.50, .01]	-.12 [-.38, .16]	.02 [-.25, .30]	**.79 [.66, .87]**	**.55 [.32, .71]**					
7. DASS Stress	-.09 [-.36, .18]	-.01 [-.28, .26]	.06 [-.22, .33]	**.92 [.87, .95]**	**.75 [.60, .85]**	**.63 [.43, .77]**				
8. TAS Total	-.20 [-.45, .08]	**-.29 [-.52, -.02]**	-.09 [-.36, .19]	**.42 [.16, .62]**	**.47 [.22, .66]**	**.48 [.23, .67]**	.19 [-.09, .44]			
9. BPAQ Total	-.06 [-.33, .22]	-.14 [-.40, .14]	.14 [-.14, .40]	**.56 [.34, .73]**	**.56 [.34, .73]**	**.42 [.16, .62]**	**.49 [.25, .68]**	**.45 [.20, .65]**		
10. EBT Sad (balancepoint)	-.03 [-.30, .25]	-.18 [-.43, .10]	.17 [-.11, .42]	**.35 [.08, .57]**	**.32 [.05, .55]**	**.31 [.04, .53]**	**.29 [.02, .52]**	.19 [-.09, .44]	.10 [-.18, .37]	
11. EBT Angry (balancepoint)	.25 [-.02, .49]	-.17 [-.42, .11]	**.27 [.00, .51]**	.21 [-.07, .45]	.26 [-.01, .50]	.06 [-.22, .33]	.19 [-.09, .44]	.14 [-.14, .40]	.01 [-.27, .29]	**.51 [.27, .69]**

Note: ERT is Emotion Recognition Task. GFMT is Glasgow Face Matching Test. TAS is Toronto Alexithymia Scale. BPAQ is Buss Perry Aggression Questionnaire. DASS is Depression Anxiety and Stress Scale. EBT is Emotion Bias Task. Values in square brackets indicate 95% confidence interval. In bold the values with evidence for a correlation.

8.10 Appendix D (Chapter 4) – Sensitivity analysis for primary analysis

Sensitivity analysis for the primary analysis excluding two TBI participants who failed the online attention check, as well as their corresponding matched controls.

Table 8-8

Regression models for overall emotion recognition on the Bristol ERT (sensitivity)

Predictor	*b*	95% CI		*t*	*p*	Model Fit
		LL	UL			
Model 1						n = 48
(Intercept)	0.677	0.639	0.714			R^2 = 0.335
TBI	-0.127	-0.180	-0.074	-4.82	<.001	95% CI [0.12, 0.50]
Model 2						
(Intercept)	0.684	0.642	0.726			n= 48
TBI	-0.121	-0.177	-0.065	-4.34	<.001	R^2 = 0. 344
DASS Anxiety score	-0.002	-0.006	0.003	-0.76	0.45	95% CI [0.11, 0.50]
Model 3						
(Intercept)	0.422	0.167	0.676			
TBI	-0.099	-0.156	-0.042	-3.52	0.001	
DASS Anxiety score	-0.003	-0.008	0.001	-1.49	0.14	
Sex (Female)	0.002	-0.052	0.057	0.09	0.93	
Age left education	-0.004	-0.011	0.002	-1.39	0.17	n = 47*
GFMT (total hits)	0.009	0.004	0.015	3.56	0.001	R^2 = 0.522
TAS (total score)	0.001	-0.002	0.004	0.70	0.49	95% CI [0.22, 0.62]

Note: TBI is Traumatic Brain Injury. DASS is Depression, Anxiety, and Stress Scale. GFMT is Glasgow Face Matching Test. TAS is Toronto Alexithymia Scale. b is the unstandardised regression coefficient from a multiple linear regression. LL is the lower limit and UL the upper limit of the 95% Confidence Interval. t is the t-test coefficient.
** One patient did not complete the TAS so was dropped at this stage in the analysis.*
Model 1 - Association between TBI and overall emotion recognition measured on the Bristol ERT
Model 2 - Model 1 adjusted for anxiety using the DASS anxiety subscale score
Model 3 - Model 2 additionally adjusted for sex, age left education, GFMT, and TAS

8.11 Appendix E (Chapter 4) – Association between TBI and raw hits and false alarms for each individual emotion on the Bristol ERT

Table 8-9

Results from regression analysis investigating association between TBI and emotion recognition performance in terms of hits and false alarms for each emotion on the Bristol ERT

	Model 1 - Unadjusted					Model 2 - Adjusted for Anxiety					Model 3 - Fully Adjusted				
Emotion	***b***	**95% CI**		***t***	***p***	***b***	**95% CI**		***t***	***p***	***b***	**95% CI**		***t***	***p***
		LL	**UL**				**LL**	**UL**				**LL**	**UL**		
HITS (number of correct responses for each emotion)															
Angry	-1.81	-3.30	-0.31	-2.43	0.02	-1.55	-3.09	-0.004	-2.01	0.05	-1.51	-3.24	0.23	-1.75	0.09
Disgust	-2.65	-4.70	-0.61	-2.61	0.01	-2.771	-4.91	-0.63	-2.6	0.01	-2.11	-4.39	0.17	-1.87	0.07
Fear	-2.62	-4.32	-0.92	-3.09	<.001	-2.26	-4.01	-0.52	-2.61	0.01	-1.49	-3.32	0.34	-1.64	0.11
Happy	-0.58	-2.19	1.04	-0.72	0.48	-0.24	-1.90	1.42	-0.29	0.77	0.13	-1.57	1.84	0.16	0.88
Sad	-3.27	-4.41	-2.13	-5.78	<.001	-3.16	-4.35	-1.98	-5.35	<.001	-3.00	-4.33	-1.66	-4.52	<.001
Surprise	-0.46	-1.29	0.37	-1.12	0.27	-0.75	-1.56	0.07	-1.85	0.07	-0.49	-1.36	0.39	-1.12	0.27
FALSE ALARMS (number of misidentifications made for each emotion)															
Angry	2.12	0.80	3.43	3.23	<.001	2.03	0.65	3.40	2.96	<.001	1.44	-0.04	2.92	1.97	0.06
Disgust	1.27	-0.46	3.00	1.48	0.15	1.08	-0.72	2.88	1.2	0.23	0.93	-1.02	2.89	0.96	0.34
Fear	3.39	0.91	5.86	2.74	0.01	3.26	0.66	5.86	2.52	0.02	2.50	-0.17	5.16	1.89	0.07
Happy	-0.19	-2.66	2.27	-0.16	0.88	0.09	-2.48	2.66	0.07	0.95	1.03	-1.64	3.71	0.78	0.44
Sad	-0.85	-3.69	2.00	-0.6	0.55	-0.97	-3.95	2.01	-0.66	0.51	-1.44	-4.73	1.86	-0.88	0.38
Surprise	5.65	3.02	8.29	4.31	<.001	5.25	2.52	7.98	3.86	<.001	3.99	1.26	6.71	2.94	0.01

Note: b is the TBI unstandardised regression coefficient from a multiple linear regression. LL is the lower limit and UL the upper limit of the 95% Confidence Interval. t is the t-test coefficient. Model 1 - Association between TBI and emotion specific outcomes on the Bristol ERT. Model 2 - Model 1 adjusted for anxiety on the DASS. Model 3 - Model 2 additionally adjusted for sex, age left education, GFMT, and TAS

8.12 Appendix F (Chapter 4) – Anxiety estimates from regression analysis for emotion specific associations with TBI and anxiety

Table 8-10

Association between anxiety and emotion recognition performance for each emotion on the Bristol ERT based on model 2 of the regression analysis

Anxiety estimates from regression model 2					
Emotion	***b***	**95% CI**		***t***	***p***
		LL	**UL**		
UNBIASED HIT RATE (HU)					
Angry	0.00	-0.002	0.001	-0.98	0.33
Disgust	0.00	-0.001	0.001	-0.14	0.89
Fear	-0.001	-0.002	0.001	-0.82	0.42
Happy	0.00	0.00	0.001	0.89	0.38
Sad	0.00	-0.001	0.001	-0.78	0.44
Surprise	0.00	-0.001	0.00	-0.81	0.42
BIAS SCORE (number of responses given for each emotion)					
Angry	-0.05	-0.21	0.11	-0.63	0.53
Disgust	0.09	-0.12	0.30	0.84	0.41
Fear	-0.07	-0.28	0.15	-0.60	0.55
Happy	-0.18	-0.47	0.11	-1.22	0.23
Sad	0.01	-0.27	0.28	0.04	0.97
Surprise	0.20	-0.03	0.43	1.75	0.09

Note: b is the anxiety unstandardised regression coefficient from a multiple linear regression. LL is the lower limit and UL the upper limit of the 95% Confidence Interval. t is the t-test coefficient.

Model 2 - Association between TBI and emotion specific outcomes on the Bristol ERT adjusted for anxiety on the DASS

8.13 Appendix A (Chapter 5) - Prolific screeners

10. **Age**
    What is your date of birth?
    Setting: Minimum 18 years - Maximum 100 years

11. **Current Country of Residence**
    In what country do you currently reside?
    Response required: United Kingdom

12. **Fluent languages**
    Which of the following languages are you fluent in?
    Response required: English

13. **Vision**
    Do you have normal or corrected-to-normal vision? (i.e. You can see colour normally, and if you need glasses, you are wearing them or contact lenses)
    Response required*:* Yes

14. **Mild Cognitive Impairment/Dementia**
    Have you ever been diagnosed with mild cognitive impairment or dementia?
    Response required: No

15. **Autism Spectrum Disorder**
    Have you received a formal clinical diagnosis of autistic spectrum disorder, made by a psychiatrist, psychologist, or other qualified medical specialist? This includes Asperger's syndrome, Autistic Disorder, High Functioning Autism or Pervasive Developmental Disorder
    Response required: No

16. **Chronic Disease**
    Have you been diagnosed with any chronic diseases such as diabetes, heart disease, stroke etc.?
    Response required*:* No

17. **Multiple sclerosis**
    Have you ever been diagnosed with multiple sclerosis (MS)?
    Response required: No

18. **Head Injury**
    Have you ever had an injury to the head that's caused you to be knocked out and/or dazed and confused for a period of time (E.g. from a fall, blow to the head, road traffic accident)?
    Response required for TBI group: Yes
    Response required for control group: No

19. **Head Injury: Knock out history**
    Have you ever had an injury to the head that's caused you to be knocked out for a period of time (E.g. from a fall, blow to the head, road traffic accident)?
    Response required for TBI group: Yes
    Response required for control group: No

8.14 Appendix B (Chapter 5) – Correlations between variables included in this study

Table 8-11

Pearson correlation matrix for primary outcomes of the variables included in this study

Variable	1	2	3	4	5	6	7	8	9	10	11
1. Bristol ERT (total hits)											
2. Age	**-.35**										
	[-.51, -.18]										
3. Education	.12	-.06									
	[-.07, .30]	[-.25, .13]									
4. GFMT	**.31**	-.10	.06								
(total hits)	**[.13, .47]**	[-.29, .08]	[-.12, .25]								
5. DASS Total	.09	-.15	-.04	-.10							
	[-.09, .28]	[-.33, .03]	[-.23, .14]	[-.28, .09]							
6. DASS Depression	.13	-.08	-.04	-.05	**.91**						
	[-.05, .31]	[-.26, .11]	[-.22, .15]	[-.23, .14]	**[.88, .94]**						
7. DASS Anxiety	.03	**-.26**	-.05	-.15	**.91**	**.74**					
	[-.16, .21]	**[-.42, -.07]**	[-.23, .14]	[-.33, .03]	**[.87, .94]**	**[.64, .81]**					
8. DASS Stress	.08	-.12	-.04	-.10	**.93**	**.75**	**.81**				
	[-.11, .26]	[-.30, .07]	[-.23, .15]	[-.28, .09]	**[.90, .95]**	**[.65, .82]**	**[.74, .87]**				
9. TAS Total	-.14	-.13	-.07	**-.24**	**.57**	**.47**	**.50**	**.59**			
	[-.32, .05]	[-.31, .06]	[-.25, .12]	**[-.41, -.06]**	**[.42, .68]**	**[.30, .60]**	**[.35, .63]**	**[.45, .70]**			
10. BPAQ Total	.02	-.16	.05	-.09	**.50**	**.40**	**.41**	**.56**	**.54**		
	[-.17, .21]	[-.34, .03]	[-.14, .24]	[-.28, .10]	**[.35, .63]**	**[.23, .55]**	**[.24, .56]**	**[.41, .67]**	**[.39, .66]**		
11. EBT Sad	-.18	.12	-.04	-.03	.14	.06	.15	**.19**	.10	.08	
(balance point)	[-.35, .01]	[-.07, .30]	[-.22, .15]	[-.21, .16]	[-.04, .32]	[-.12, .25]	[-.04, .32]	**[.01, .37]**	[-.09, .28]	[-.11, .26]	
12. EBT Angry	-.17	.04	-.04	-.07	-.09	-.11	-.11	-.05	.06	.11	.18
(balance point)	[-.35, .01]	[-.15, .23]	[-.22, .15]	[-.26, .12]	[-.28, .10]	[-.29, .08]	[-.29, .08]	[-.23, .14]	[-.13, .25]	[-.08, .29]	[-.01, .35]

Note: ERT is Emotion Recognition Task. GFMT is Glasgow Face Matching Test. TAS is Toronto Alexithymia Scale. BPAQ is Buss Perry Aggression Questionnaire. DASS is Depression Anxiety and Stress Scale. Values in square brackets indicate the 95% confidence interval for each correlation. In bold the values with evidence for a correlation.

8.15 Appendix C (Chapter 5) – Sensitivity analysis

Participants who reported having loss of consciousness longer than 30 minutes or were unsure about how long they were unconscious were removed from the TBI group for this analysis, so 47 participants with mild TBI and 55 controls were included.

Table 8-12

Regression models for overall emotion recognition on the Bristol ERT

Predictor	*b*	95% CI		*t*	*p*	Model Fit
		LL	UL			
Model 1						n = 102
(Intercept)	0.68	0.660	0.704			R^2 = .039
Mild TBI	-0.03	-0.065	-0.001	-2.01	0.05	95% CI [.00,.13]
Model 2						
(Intercept)	0.675	0.650	0.701			n = 102
Mild TBI	-0.038	-0.072	-0.004	-2.24	0.03	R^2 = .050
DASS Anxiety score	0.001	-0.001	0.003	1.62	0.29	95% CI [.00,.14]
Model 3						
(Intercept)	0.716	0.627	0.806			
Mild TBI	-0.021	-0.057	0.016	-1.12	0.26	
DASS Anxiety score	0.000	-0.002	0.002	0.08	0.93	
Sex (Female)	-0.017	-0.052	0.018	-0.94	0.35	n = 102
Age	-0.002	-0.003	-0.001	-3.41	<.001	R^2 = .179
Age left education	0.002	-0.001	0.005	1.18	0.24	95% CI [.03,.28]
Model 4						
(Intercept)	0.615	0.455	0.775			
Mild TBI	-0.019	-0.055	0.017	-1.05	0.30	
DASS Anxiety score	0.001	-0.001	0.004	0.91	0.37	
Sex (Female)	-0.008	-0.043	0.027	-0.48	0.63	
Age	-0.002	-0.003	-0.001	-3.08	0.002	
Age left education	0.002	-0.002	0.005	0.96	0.34	n = 102
GFMT (total hits)	0.004	0.001	0.007	2.44	0.02	R^2 = .245
TAS (total score)	-0.001	-0.002	0.001	-1.06	0.29	95% CI [.06,.33]

Note: TBI is Traumatic Brain Injury. DASS is Depression, Anxiety, and Stress Scale. GFMT is Glasgow Face Matching Test. TAS is Toronto Alexithymia Scale. b is the unstandardised regression coefficient from a multiple linear regression. LL is the lower limit and UL the upper limit of the 95% Confidence Interval. t is the t-test coefficient.

Model 1 - Association between TBI and overall emotion recognition on the Bristol ERT
Model 2 - Model 1 adjusted for anxiety using the DASS anxiety subscale score
Model 3 - Model 2 additionally adjusted for sex, age, and age left education
Model 4 – Model 3 additionally adjusted for total hits on the GFMT and TAS score

8.16 Appendix D (Chapter 5) – Anxiety estimates from regression analysis for emotion specific associations with TBI and anxiety

Table 8-13

Association between anxiety and emotion recognition performance for each emotion on the Bristol ERT based on model 2 of the regression analysis

Anxiety estimates from regression model 2					
Emotion	***b***	**95% CI**		***t***	***p***
		LL	**UL**		
UNBIASED HIT RATE (HU)					
Angry	0.00	0.00	0.001	0.55	0.59
Disgust	0.00	-0.001	0.00	-1.29	0.20
Fear	0.001	0.00	0.002	1.74	0.08
Happy	0.00	0.00	0.001	0.61	0.54
Sad	0.00	0.00	0.001	1.51	0.13
Surprise	0.00	0.00	0.00	0.03	0.98
BIAS SCORE (number of responses given for each emotion)					
Angry	0.06	-0.05	0.16	1.03	0.3
Disgust	0.12	-0.03	0.27	1.57	0.12
Fear	-0.03	-0.19	0.12	-0.39	0.7
Happy	-0.08	-0.24	0.08	-0.98	0.33
Sad	-0.02	-0.15	0.1	-0.37	0.71
Surprise	-0.04	-0.19	0.11	-0.55	0.59

Note: b is the anxiety unstandardised regression coefficient from a multiple linear regression. LL is the lower limit and UL the upper limit of the 95% Confidence Interval. t is the t-test coefficient.

Model 2 - Association between TBI and emotion specific outcomes on the Bristol ERT adjusted for anxiety on the DASS

8.17 Appendix A (Chapter 6) - Number of participants included in each analysis

Table 8-14

Number of participants in each injury group that completed the Bristol ERT at age 24

Time of Injury	Unadjusted			Model 1			Model 2		
	No Injury	mTBI	Broken Bones	No Injury	mTBI	Broken Bones	No Injury	mTBI	Broken Bones
Age 17 to 25	748	254	144	629	200	111	616	194	108
Birth to age 25	612	530	761	515	434	620	506	425	607

Note: mTBI is mild Traumatic Brain Injury. Unadjusted: Number of participants in each injury group who also completed the Bristol ERT. Model 1: As Unadjusted after exclusions due to missing information about age at 24 clinic, sex at birth, and maternal social class covariates have been applied. Model 2: As Model 1 after exclusions due to missing data for the working memory covariate have been applied.

Table 8-15

Number of participants with Generalised Anxiety Disorder in the injury groups

Time of Injury	Anxiety	Unadjusted			Model 1		
		No Injury	mTBI	Broken Bones	No Injury	mTBI	Broken Bones
Age 17 to 25	No anxiety	690	219	144	580	178	112
	GAD	63	36	9	49	26	6
Birth to age 25	No anxiety	566	470	717	475	390	587
	GAD	54	74	53	41	56	44

Notes: GAD is Generalised Anxiety Disorder and mTBI is mild Traumatic Brain Injury. Unadjusted: Number of participants in each injury group for whom GAD at age 24 is available. Model 1: As Unadjusted after exclusions due to missing information about age at 24 clinic, sex at birth, and maternal social class covariates have been applied.

Table 8-16

Number of participants with or without anxiety that competed the Bristol ERT

Anxiety	Unadjusted		Model 1		Model 2	
	No	Yes	No	Yes	No	Yes
GAD	2,957	337	2,375	261	2,332	258
Anxiety in past week	2,337	1,121	1,892	865	1,857	851

Note: GAD is Generalised Anxiety Disorder. Unadjusted: Number of participants with and without anxiety who completed the Bristol ERT. Model 1: As Unadjusted after exclusions due to missing information about age at 24 clinic, sex at birth, and maternal social class covariates have been applied. Model 2: As Model 1 after exclusions due to missing data for the working memory covariate has been applied.

8.18 Appendix B (Chapter 6) - Association between adult injury and hits for each emotion

Table 8-17

Results from multiple linear regression analysis investigating impact of injury between ages 17 and 25 on raw hits for each emotion

Time of Injury		Unadjusted					Model 1					Model 2				
Age 17 to 25		n	Estimate	95% CI		p	n	Estimate	95% CI		p	n	Estimate	95% CI		p
Emotion	Injury			LL	UL				LL	UL				LL	UL	
Angry	mTBI vs None	1,146	0.228	-0.123	0.580	0.20	940	0.226	-0.166	0.618	0.26	831	0.416	0.002	0.830	0.05
	BB vs None		-0.271	-0.711	0.170	0.23		-0.197	-0.693	0.300	0.44		-0.245	-0.798	0.308	0.39
	mTBI vs BB		0.499	-0.006	1.004	0.05		0.423	-0.142	0.988	0.14		0.660	0.035	1.286	0.04
Disgust	mTBI vs None		-0.041	-0.357	0.275	0.80		-0.017	-0.370	0.336	0.92		0.021	-0.359	0.402	0.91
	BB vs None		-0.129	-0.525	0.266	0.52		0.077	-0.370	0.524	0.74		-0.122	-0.630	0.386	0.64
	mTBI vs BB		0.088	-0.366	0.542	0.70		-0.094	-0.604	0.415	0.72		0.143	-0.432	0.718	0.63
Fear	mTBI vs None		-0.251	-0.731	0.230	0.31		-0.219	-0.760	0.323	0.43		-0.284	-0.860	0.293	0.33
	BB vs None		-0.705	-1.307	-0.104	0.02		-0.728	-1.414	-0.041	0.04		-0.545	-1.315	0.224	0.16
	mTBI vs BB		0.455	-0.235	1.145	0.20		0.509	-0.273	1.290	0.20		0.262	-0.609	1.132	0.56
Happy	mTBI vs None		0.028	-0.278	0.334	0.86		0.190	-0.152	0.531	0.28		0.130	-0.242	0.502	0.49
	BB vs None		0.152	-0.231	0.536	0.44		0.225	-0.207	0.658	0.31		0.293	-0.204	0.789	0.25
	mTBI vs BB		-0.125	-0.564	0.315	0.58		-0.035	-0.528	0.457	0.89		-0.163	-0.725	0.399	0.57
Sad	mTBI vs None		-0.135	-0.425	0.156	0.36		-0.155	-0.483	0.174	0.36		-0.212	-0.555	0.131	0.23
	BB vs None		-0.152	-0.516	0.212	0.41		-0.055	-0.471	0.361	0.80		-0.057	-0.515	0.401	0.81
	mTBI vs BB		0.017	-0.400	0.435	0.94		-0.100	-0.573	0.374	0.68		-0.155	-0.674	0.363	0.56
Surprise	mTBI vs None		-0.188	-0.406	0.029	0.09		-0.113	-0.360	0.135	0.37		-0.132	-0.400	0.135	0.33
	BB vs None		-0.389	-0.662	-0.117	0.01		-0.383	-0.696	-0.070	0.02		-0.386	-0.743	-0.029	0.03
	mTBI vs BB		0.201	-0.112	0.513	0.21		0.271	-0.086	0.627	0.14		0.254	-0.150	0.658	0.22

Note: Sample size reduces as participants who are missing covariate data used in that model are excluded. mTBI refers to mild Traumatic Brain Injury. BB refers to Broken Bones. None refers to No Injury. Estimate is regression coefficient from a multiple linear regression.

8.19 Appendix C (Chapter 6) - Sensitivity analysis for Step 1. Results tables replicated after outliers on the Bristol ERT were removed

Table 8-18

Results from regression analysis investigating impact of injury on overall emotion recognition with outliers removed

Time of Injury		Unadjusted					Model 1					Model 2				
		n	Estimate	95% CI		p	n	Estimate	95% CI		p	n	Estimate	95% CI		p
				LL	UL				LL	UL				LL	UL	
Age 17 to 25	mTBI vs None	1,135	-0.005	-0.016	0.006	0.41	933	-0.001	-0.013	0.011	0.81	911	-0.001	-0.012	0.011	0.93
	BB vs None		-0.012	-0.026	0.002	0.09		-0.011	-0.026	0.005	0.18		-0.008	-0.023	0.007	0.29
	mTBI vs BB		0.007	-0.008	0.023	0.36		0.009	-0.008	0.026	0.31		0.008	-0.010	0.025	0.39
Birth to age 25	mTBI vs None	1,881	-0.007	-0.016	0.002	0.10	1,553	-0.002	-0.011	0.008	0.71	1,524	-0.001	-0.010	0.009	0.90
	BB vs None		-0.006	-0.014	0.003	0.18		-0.003	-0.012	0.006	0.51		-0.001	-0.010	0.008	0.82
	mTBI vs BB		-0.002	-0.010	0.007	0.68		0.001	-0.008	0.010	0.81		0.000	-0.009	0.009	0.93

Note: Sample size reduces for each model as participants who are missing covariate data used in that model are excluded. mTBI refers to mild Traumatic Brain Injury. BB refers to Broken Bones. None refers to No Injury. Estimate is regression coefficient from a multiple linear regression.
Models:
Unadjusted: Impact of receiving an injury on emotion recognition performance measured using the Bristol ERT at age 24
Model 1: As Unadjusted and additionally adjusted for age at 24 clinic, sex at birth, and maternal social class
Model 2: As Model 1 and additionally adjusted for working memory (2-back) at age 24

Table 8-19

Results from regression analysis investigating impact of injury on unbiased hit rate for each emotion with outliers removed

Time of Injury		Unadjusted					Model 1					Model 2				
Age 17 to 25		n	Estimate	95% CI		p	n	Estimate	95% CI		p	n	Estimate	95% CI		p
Emotion	Injury			LL	UL				LL	UL				LL	UL	
Angry	mTBI vs None	1,135	-0.002	-0.005	0.000	0.05	933	-0.001	-0.004	0.001	0.30	824	-0.002	-0.005	0.001	0.19
	BB vs None		-0.001	-0.004	0.002	0.67		0.000	-0.003	0.004	0.95		0.000	-0.004	0.004	0.87
	mTBI vs BB		-0.002	-0.005	0.002	0.34		-0.002	-0.005	0.002	0.44		-0.002	-0.006	0.003	0.47
Disgust	mTBI vs None		0.002	-0.001	0.004	0.20		0.001	-0.001	0.004	0.29		0.002	-0.001	0.005	0.26
	BB vs None		0.000	-0.003	0.003	0.82		-0.001	-0.004	0.003	0.75		-0.001	-0.005	0.003	0.72
	mTBI vs BB		0.001	-0.002	0.005	0.49		0.002	-0.002	0.006	0.31		0.002	-0.002	0.007	0.28
Fear	mTBI vs None		-0.001	-0.006	0.003	0.49		-0.002	-0.006	0.003	0.44		-0.003	-0.008	0.002	0.21
	BB vs None		-0.004	-0.010	0.001	0.09		-0.004	-0.010	0.002	0.18		-0.005	-0.011	0.002	0.15
	mTBI vs BB		0.003	-0.003	0.009	0.31		0.002	-0.004	0.009	0.52		0.002	-0.006	0.009	0.64
Happy	mTBI vs None		0.000	-0.002	0.002	0.90		0.000	-0.002	0.003	0.87		0.001	-0.002	0.004	0.45
	BB vs None		-0.003	-0.006	0.000	0.07		-0.002	-0.006	0.001	0.22		-0.003	-0.006	0.001	0.19
	mTBI vs BB		0.003	-0.001	0.006	0.14		0.002	-0.002	0.006	0.23		0.004	-0.001	0.008	0.10
Sad	mTBI vs None		0.000	-0.003	0.002	0.71		0.000	-0.003	0.003	0.83		0.000	-0.003	0.003	0.84
	BB vs None		-0.002	-0.006	0.001	0.15		-0.002	-0.005	0.002	0.33		-0.001	-0.005	0.003	0.58
	mTBI vs BB		0.002	-0.002	0.006	0.31		0.002	-0.003	0.006	0.48		0.001	-0.004	0.006	0.72
Surprise	mTBI vs None		-0.002	-0.004	0.000	0.06		-0.002	-0.004	0.001	0.22		-0.002	-0.005	0.001	0.17
	BB vs None		-0.002	-0.005	0.001	0.15		-0.003	-0.006	0.001	0.10		-0.001	-0.005	0.002	0.44
	mTBI vs BB		0.000	-0.003	0.003	0.96		0.001	-0.003	0.005	0.56		0.000	-0.005	0.004	0.83

Note: Sample size reduces for each model as participants who are missing covariate data used in that model are excluded. mTBI refers to mild Traumatic Brain Injury. BB refers to Broken Bones. None refers to No Injury. Estimate is regression coefficient from a multiple linear regression. Unadjusted: Impact of receiving an injury on emotion recognition performance measured using the Bristol ERT at age 24. Model 1: As Unadjusted and additionally adjusted for age at 24 clinic, sex at birth, and maternal social class Model 2: As Model 1 and additionally adjusted for working memory (2-back) at age 24.

Table 8-20

Results from regression analysis investigating impact of injury on number of responses given for each emotion with outliers removed

Time of Injury		Unadjusted					Model 1					Model 2				
Age 17 to 25		n	Estimate	95% CI		p	n	Estimate	95% CI		p	n	Estimate	95% CI		p
Emotion	Injury			LL	UL				LL	UL				LL	UL	
Angry	mTBI vs None	1,135	0.659	0.189	1.128	0.01	933	0.507	-0.021	1.035	0.06	824	0.794	0.239	1.350	0.01
	BB vs None		-0.110	-0.702	0.482	0.72		-0.232	-0.902	0.438	0.50		-0.248	-0.992	0.496	0.51
	mTBI vs BB		0.769	0.091	1.446	0.03		0.739	-0.023	1.501	0.06		1.042	0.201	1.883	0.02
Disgust	mTBI vs None		-0.403	-0.973	0.167	0.17		-0.285	-0.931	0.362	0.39		-0.286	-0.975	0.403	0.42
	BB vs None		-0.218	-0.937	0.501	0.55		0.134	-0.687	0.955	0.75		-0.120	-1.043	0.803	0.80
	mTBI vs BB		-0.185	-1.008	0.638	0.66		-0.418	-1.352	0.515	0.38		-0.166	-1.209	0.877	0.75
Fear	mTBI vs None		-0.105	-0.729	0.518	0.74		-0.067	-0.786	0.651	0.85		-0.044	-0.824	0.736	0.91
	BB vs None		-0.423	-1.209	0.364	0.29		-0.595	-1.508	0.317	0.20		-0.200	-1.245	0.845	0.71
	mTBI vs BB		0.317	-0.582	1.217	0.49		0.528	-0.509	1.566	0.32		0.156	-1.025	1.337	0.80
Happy	mTBI vs None		-0.147	-0.985	0.691	0.73		-0.054	-1.007	0.899	0.91		-0.327	-1.361	0.707	0.54
	BB vs None		0.890	-0.167	1.947	0.10		0.686	-0.524	1.896	0.27		0.863	-0.522	2.248	0.22
	mTBI vs BB		-1.037	-2.247	0.172	0.09		-0.740	-2.116	0.637	0.29		-1.190	-2.755	0.376	0.14
Sad	mTBI vs None		-0.163	-0.821	0.496	0.63		-0.245	-1.003	0.514	0.53		-0.338	-1.155	0.480	0.42
	BB vs None		0.085	-0.746	0.916	0.84		0.122	-0.842	1.085	0.80		0.073	-1.023	1.168	0.90
	mTBI vs BB		-0.248	-1.198	0.703	0.61		-0.366	-1.462	0.729	0.51		-0.410	-1.648	0.828	0.52
Surprise	mTBI vs None		0.159	-0.444	0.763	0.60		0.143	-0.546	0.833	0.68		0.200	-0.551	0.951	0.60
	BB vs None		-0.224	-0.985	0.537	0.56		-0.114	-0.989	0.761	0.80		-0.368	-1.374	0.638	0.47
	mTBI vs BB		0.384	-0.487	1.254	0.39		0.257	-0.738	1.252	0.61		0.568	-0.569	1.705	0.33

Note: Sample size reduces for each model as participants who are missing covariate data used in that model are excluded. mTBI refers to mild Traumatic Brain Injury. BB refers to Broken Bones. None refers to No Injury. Estimate is regression coefficient from a multiple linear regression. Unadjusted: Impact of receiving an injury on emotion recognition performance measured using the Bristol ERT at age 24. Model 1: As Unadjusted and additionally adjusted for age at 24 clinic, sex at birth, and maternal social class Model 2: As Model 1 and additionally adjusted for working memory (2-back) at age 24.

8.20 Appendix D (Chapter 6) - Regression coefficients for the association between injury and presence of GAD at age 24.

Table 8-21

Results from regression analysis investigating impact of injury on presence of Generalised Anxiety Disorder at age 24

Time of Injury	Unadjusted					Model 1					Model 2				
	n	Estimate	95% CI		p	n	Estimate	95% CI		p	n	Estimate	95% CI		p
			LL	UL				LL	UL				LL	UL	
Age 17 to 25	1,161	0.588	0.143	1.020	0.01	951	0.647	0.118	1.156	0.01	778	0.295	-0.405	0.943	0.39
Birth to age 25	1,934	0.501	0.132	0.876	0.01	1,593	0.630	0.196	1.071	<0.01	1,221	-	-	-	-

Note: Sample size reduces for each model as participants who are missing covariate data used in that model are excluded. Estimate is the coefficients from a logistic regression with Generalised Anxiety Disorder at 24 as the outcome. LL is the lower limit and UL the upper limit of the 95% Confidence Interval

Models:
Unadjusted: Odds of having Generalised Anxiety Disorder at age 24 after mild TBI compared to participants without an injury.
Model 1: As Unadjusted and additionally adjusted for age at 24 clinic, sex at birth, and maternal social class
Model 2: As Model 1 and additionally adjusted for presence of Generalised Anxiety Disorder at age 17 years and 6 months

8.21 Appendix E (Chapter 6) - Association between anxiety and hits for each emotion

Table 8-22

Results from regression analysis investigating impact of anxiety on raw hits for each emotion

		Unadjusted					Model 1					Model 2				
Anxiety	**Emotion**	**n**	**Estimate**	**95% CI**		**p**	**n**	**Estimate**	**95% CI**		**p**	**n**	**Estimate**	**95% CI**		**p**
				LL	**UL**				**LL**	**UL**				**LL**	**UL**	
Generalised	Angry	3,294	0.120	-0.160	0.400	0.40	2,636	0.068	-0.249	0.385	0.67	2,590	0.083	-0.230	0.397	0.60
Anxiety	Disgust		0.107	-0.153	0.367	0.42		0.099	-0.196	0.394	0.51		0.112	-0.177	0.401	0.45
Disorder	Fear		0.125	-0.257	0.507	0.52		0.076	-0.356	0.507	0.73		0.150	-0.277	0.578	0.49
	Happy		-0.313	-0.554	-0.073	0.01		-0.387	-0.659	-0.115	0.01		-0.381	-0.654	-0.109	0.01
	Sad		0.197	-0.044	0.437	0.11		0.225	-0.048	0.498	0.11		0.244	-0.027	0.516	0.08
	Surprise		0.059	-0.115	0.233	0.50		-0.014	-0.211	0.184	0.89		-0.001	-0.199	0.198	1.00
Anxiety in the	Angry	3,458	0.255	0.079	0.431	<0.01	2,757	0.251	0.052	0.450	0.01	2,708	0.219	0.021	0.416	0.03
past week	Disgust		0.285	0.122	0.448	<0.01		0.249	0.064	0.434	0.01		0.246	0.063	0.429	0.01
	Fear		0.348	0.106	0.589	<0.01		0.322	0.048	0.596	0.02		0.303	0.032	0.575	0.03
	Happy		-0.042	-0.194	0.111	0.59		-0.091	-0.263	0.082	0.30		-0.086	-0.259	0.087	0.33
	Sad		0.201	0.049	0.353	0.01		0.210	0.038	0.383	0.02		0.209	0.037	0.381	0.02
	Surprise		0.018	-0.092	0.128	0.75		-0.040	-0.165	0.085	0.53		-0.036	-0.162	0.089	0.57

Note: Sample size reduces for each model as participants who are missing covariate data used in that model are excluded. Estimate is regression coefficient from a multiple linear regression. LL is the lower limit and UL the upper limit of the 95% Confidence Interval. Unadjusted: Impact of having anxiety on emotion recognition performance measured using the Bristol ERT at age 24. Model 1: As Unadjusted and additionally adjusted for age at 24 clinic, sex at birth, and maternal social class. Model 2: As Model 1 and additionally adjusted for working memory (n-back) at age 24.