THE EFFECT OF ACQUIRED BRAIN INJURY
ON THEORY OF MIND AND DECISION MAKING

Clare Barker-Ellis

May 2017

A THESIS SUBMITTED TO THE UNIVERSITY OF BIRMINGHAM
FOR THE DEGREE OF DOCTOR OF CLINICAL PSYCHOLOGY

Department of Clinical Psychology
School of Psychology
The University of Birmingham
This unpublished thesis/dissertation is copyright of the author and/or third parties. The intellectual property rights of the author or third parties in respect of this work are as defined by The Copyright Designs and Patents Act 1988 or as modified by any successor legislation.

Any use made of information contained in this thesis/dissertation must be in accordance with that legislation and must be properly acknowledged. Further distribution or reproduction in any format is prohibited without the permission of the copyright holder.
Overview

This thesis is submitted in partial fulfilment of the requirements of the degree of Doctor of Clinical Psychology (D.Clin.Psy). The thesis consists of two volumes. Volume One is the research component, which includes a systematic literature review, an empirical study and a public dissemination document. Volume Two is the clinical component and consists of five clinical practice reports.

VOLUME ONE

The first paper in Volume One is a meta-analysis of theory of mind impairment in individuals who have experienced a Traumatic Brain Injury (TBI). The second paper is an empirical research study of the role of cognitive biases in decision-making in individuals with a diagnosis of Multiple Sclerosis (MS). The third paper is an executive summary report for the dissemination of findings to the public and relevant stakeholders.

VOLUME TWO

The first report in Volume Two is a case formulation of a 21 year old with ulcerative colitis, low mood and anxiety, from Cognitive Behavioural and Psychodynamic perspectives. The second report is an audit of the extent to which a Physical Health Psychology Service met the standards of the NICE Clinical Guidance for depression in adults with a chronic physical health problem. The third report describes the use of a single case experimental design in a Cognitive Behavioural Therapy intervention for a 76 year old woman with generalised anxiety. The fourth report presents an example of leadership through detailing the design and delivery of Motivational Interviewing training for healthcare professionals. The final report presents a comprehensive neuropsychological assessment of a mild Traumatic Brain Injury.

All names and identifying features have been changed to maintain confidentiality.
Dedication

To Mum and Dad, for helping me begin my journey, and to Andrew, who walks beside me every step of the way.
Acknowledgments

I would like to acknowledge the Clinical Psychology Team at the University of Birmingham for their direction and support. In particular I would like to thank my supervisor Dr Chris Jones for his untiring enthusiasm and vast knowledge of statistics.

I would like to thank Dr John Woolmore and the Clinical Nurse Specialists at University Hospitals Birmingham NHS Foundation Trust, thank you for your ideas and for your time.

Special thanks are owed to Dr Katy Barker for proof reading in unusual circumstances and to the new Dr and Mrs Barker for ensuring I met my deadline.

Finally, I would like to thank all the patients, participants and health professionals who contributed to the development of these projects. I hope that my work will benefit them and others in similar situations.

Thank you to Mum and Dad for making me believe I could do it.

Thank you to Andrew for making sure I did.
Literature Review: Theory of Mind Impairment in Individuals who have Experienced Traumatic Brain Injury: A Meta-Analysis

Abstract ....................................................................................................................................... 1
Introduction ................................................................................................................................ 2
Aims ........................................................................................................................................... 9
Method ........................................................................................................................................ 9

Search Strategy ......................................................................................................................... 9
Quality Framework ...................................................................................................................... 15
Quality Ratings ............................................................................................................................ 18
Procedure for Quantitative Synthesis .......................................................................................... 21
Quantifying Heterogeneity ........................................................................................................... 23

Results ...................................................................................................................................... 24

First Order Belief Tasks .............................................................................................................. 26
Second Order Belief Tasks .......................................................................................................... 29
Understanding Indirect Speech Tasks .......................................................................................... 32
Faux Pas Tasks ............................................................................................................................ 35

Discussion ................................................................................................................................ 43

Conclusions ............................................................................................................................... 50

References ................................................................................................................................ 52
**Empirical Research Paper: Healthcare Decision Making: How can Cognitive Bias be Reduced to Aid Treatment Choice for Patients with Multiple Sclerosis?**

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>65</td>
</tr>
<tr>
<td>Introduction</td>
<td>66</td>
</tr>
<tr>
<td>Cognitive Biases in Decision-Making</td>
<td>66</td>
</tr>
<tr>
<td>Shared Decision-Making</td>
<td>69</td>
</tr>
<tr>
<td>Cognitive Biases in Healthcare</td>
<td>71</td>
</tr>
<tr>
<td>Multiple Sclerosis (MS)</td>
<td>73</td>
</tr>
<tr>
<td>The Present Study</td>
<td>75</td>
</tr>
<tr>
<td>Hypotheses</td>
<td>77</td>
</tr>
<tr>
<td>Method</td>
<td>78</td>
</tr>
<tr>
<td>Results</td>
<td>92</td>
</tr>
<tr>
<td>Experiment 1: Personal Relevance</td>
<td>92</td>
</tr>
<tr>
<td>Experiment 2: Very Rare Adverse Events</td>
<td>98</td>
</tr>
<tr>
<td>Experiment 3: Presentation of Risk Information</td>
<td>103</td>
</tr>
<tr>
<td>Discussion</td>
<td>105</td>
</tr>
<tr>
<td>Conclusions</td>
<td>114</td>
</tr>
<tr>
<td>References</td>
<td>115</td>
</tr>
<tr>
<td>References</td>
<td>125</td>
</tr>
<tr>
<td>Executive Summary: Public Domain Briefing Document</td>
<td></td>
</tr>
<tr>
<td>References</td>
<td>132</td>
</tr>
</tbody>
</table>
List of Appendices

Appendix A. Participant Recruitment Poster ................................................................. 135
Appendix B. Information Sheet ..................................................................................... 136
Appendix C. Consent Sheet .......................................................................................... 137
Appendix D. Debrief Sheet with Support Services and Contact Details ...................... 138
Appendix E. Operationalisation of Variables ............................................................... 139
Appendix F. Examples of Decision-Making Questions ................................................. 141
Appendix G. MS Specific Measures ............................................................................. 146
Appendix H. Expanded Disability Status Scale (EDSS; Kurtzke, 1983) (Survey Version) .. 147
Appendix I. World Health Organization (WHO) Disability Assessment Schedule 2.0, 12 Item Version (WHODAS 2.0; World Health Organization, 2010) ...................... 148
Appendix J. Patient Health Questionnaire-9 (PHQ 9; Kroenke et al., 2001) ............... 150
Appendix K. Ten-Item Personality Inventory (TIPI; Gosling, Rentfrow, & Swann, 2003) .. 151
Appendix L. MS Group Variables Data ....................................................................... 152
Appendix M. Skew and Kurtosis of the Data Set .......................................................... 153
Appendix N. Moderator Analyses for Experiment 1 ................................................... 155
Appendix O. Covariate Analyses for Experiment 1 ...................................................... 157
Appendix P. Non-Parametric Test for Experiment 1 .................................................... 158
Appendix Q. Moderator Analyses for Experiment 2 ................................................... 159
Appendix R. Covariate Analyses for Experiment 2 ..................................................... 161
Appendix S. Non-Parametric Test for Experiment 2 .................................................... 162
Appendix T. NHS Research Ethics Committee: Favourable Ethical Opinion Letter ....... 163
Appendix U. University Hospitals Birmingham Research Governance: Project Authorisation Letter ................................................................................................................. 164
List of Figures

Literature Review: Theory of Mind Impairment in Individuals who have Experienced Traumatic Brain Injury: A Meta-Analysis

Figure 1. The development of ToM abilities in childhood. ................................................................. 4

Figure 2. Flowchart of meta-analysis search strategy and results....................................................... 12

Figure 3. Forest plot of first order belief task studies................................................................. 26

Figure 4. Forest plot of influence of first order belief task studies on meta-analytic effect. ....27

Figure 5. Funnel plot (a) and trim and fill analysis (b) for first order belief task studies.......27

Figure 6. Forest plot of second order belief task studies................................................................. 29

Figure 7. Forest plot of influence of second order belief task studies on meta-analytic effect30

Figure 8. Funnel plot (a) and trim and fill analysis (b) for second order belief task studies. ..30

Figure 9. Forest plot of understanding indirect speech studies....................................................... 32

Figure 10. Forest plot of influence of understanding indirect speech task studies on meta-
analytic effect. ................................................................................................................................. 33

Figure 11. Funnel plot (a) and trim and fill analysis (b) for understanding indirect speech studies. .......................................................................................................................................... 33

Figure 12. Forest plot of faux pas task studies.................................................................................. 35

Figure 13. Forest plot of influence of faux pas task studies on meta-analytic effect.................36

Figure 14. Funnel plot (a) and trim and fill analysis (b) for faux pas task studies.................36
Empirical Research Paper: Healthcare Decision Making: How can Cognitive Bias be Reduced to Aid Treatment Choice for Patients with Multiple Sclerosis?

Figure 1. Dual process model for diagnostic reasoning applied to decision-making

(Croskerry, 2009b). .......................................................................................................... 68

Figure 2. Pie chart presenting MS group medication at time of participation. ................. 80

Figure 3. Histogram 3a: MS group ratings of severity of MS in general (for most people);

Histogram 3b: MS group ratings of severity of personal disease condition.................. 81

Figure 4. Histogram of MS group level of functioning based on EDSS score. .............. 82

Figure 5. An example question from Experiment 1, from the low severity condition, with

health risk and incongruent format................................................................................. 88

Figure 6. An example question from Experiment 2, manipulating the risk type MS and high

probability of adverse outcome..................................................................................... 88

Figure 7. Experiment 3 decision-making question presented as text.............................. 89

Figure 8. Experiment 3 decision-making question presented graphically..................... 90

Figure 9. Graphical representation of marginal means for Experiment 1....................... 95

Figure 10. Graphical representation of marginal means for Experiment 2.................... 101

Figure 11. Dual process model for diagnostic reasoning and decision-making with emotional

elements highlighted (Croskerry, 2009b). ........................................................................ 109

Executive Summary: Public Domain Briefing Document

Figure 1. The development of ToM abilities in childhood. ........................................... 126
List of Tables

Literature Review: Theory of Mind Impairment in Individuals who have Experienced Traumatic Brain Injury: A Meta-Analysis

Table 1. Exclusion and Inclusion Criteria ........................................................................................................... 10
Table 2. Effect Sizes Reported by the Studies Included in the Meta-Analysis ...................................................... 14
Table 3. Risk of Bias Domains and Items in the Quality Framework ................................................................. 16
Table 4. Risk of Bias Ratings Based on Quality Framework .................................................................................. 19
Table 5. Meta-Analysis Results Summary by Theory of Mind Task ................................................................. 25
Table 6. Results from Meta-regression Analyses of the Impact of Continuous Moderator Variables on ToM Task Effect Sizes for the Present Study and Previous Meta-analysis by Martín-Rodríguez and León-Carrión (2010) ........................................................................ 41
Table 7. Results from Subgroup Analysis of the Impact of Categorical Variable of Severity of TBI on ToM Task Effect Sizes ........................................................................................................ 42

Empirical Research Paper: Healthcare Decision Making: How can Cognitive Bias be Reduced to Aid Treatment Choice for Patients with Multiple Sclerosis?

Table 1. Characteristic Properties of the Two Systems in a Dual Process Model of Reasoning; Adapted from (Croskerry, 2005, 2009b; Dawson, 1993; Evans, 2008) ......................................................... 67
Table 2. Participant Demographic Information .................................................................................................. 79
Table 3. ANOVA Results with Greenhouse-Geisser Corrected Estimates for Experiment 1 ... 93
Table 4. Maximum Likelihood Polyserial Correlation of Highest Level of Education and ToM Tasks ............................................................................................................................................... 97
Table 5. ANOVA Results with Greenhouse-Geisser Corrected Estimates for Experiment 2 ... 99
Table 6. Participant Responses to Experiment 3 ................................................................................................. 104
Table 7. *Tests of Between-Subjects Effects in Experiment 3* .................................................. 105

*Executive Summary: Public Domain Briefing Document*

Table 1. *Characteristic Properties of the Two Systems in a Dual Process Model of Reasoning; Adapted from (Croskerry, 2005, 2009b; Dawson, 1993; Evans, 2008)* ......................... 130
VOLUME TWO

List of Contents

Clinical Practice Report 1: Case Formulation of Adam, a 21 year old with Ulcerative Colitis, Low Mood and Anxiety, from Cognitive Behavioural and Psychodynamic Perspectives

Abstract........................................................................................................................... 1

Background Information and Assessment................................................................. 2

Formulation.................................................................................................................... 14

  Cognitive Behavioural Formulation................................................................. 15

  Psychodynamic Formulation........................................................................ 19

Critical Thinking.......................................................................................................... 24

Reflective Practice......................................................................................................... 26

References..................................................................................................................... 28

Clinical Practice Report 2: An Audit of the Extent to which a Physical Health Psychology Service meets the Standards of the NICE Clinical Guideline 91: Depression in Adults with a Chronic Physical Health Problem: Treatment and Management

Abstract........................................................................................................................ 34

Introduction.................................................................................................................... 35

  Depression and Chronic Physical Health Conditions...................................... 35

  NICE Guidance and this Guideline (CG91)...................................................... 36

  The Service............................................................................................................. 38

Aim............................................................................................................................... 39
Clinical Practice Report 3: Cognitive Behavioural Therapy with a 76 year old Woman with Generalised Anxiety: A Single Case Experimental Design

Abstract.................................................................................................................. 79
Case Summary......................................................................................................... 80
Assessment.............................................................................................................. 81
  Life History....................................................................................................... 82
  Onset of Difficulties........................................................................................... 83
  Current Presentation......................................................................................... 84
  Goals and Aims................................................................................................. 85
Empirical Evidence and Theoretical Models............................................................. 85
Clinical Practice Report 4: Training Practitioners in Motivational Interviewing in a Community Learning Disability Service: An Example of Leadership

Abstract ................................................................................................................. 115

Background ............................................................................................................. 116

Policy Context .......................................................................................................... 119

Continuing Professional Development (CPD) .......................................................... 119

Theory and Research Context .................................................................................. 126

Roles and Aims ......................................................................................................... 131

The Role of the Learning Disability Service ............................................................ 131

The Role of the Psychology Team ........................................................................... 132

Aims .......................................................................................................................... 134

Work Undertaken .................................................................................................... 135
Example 1: Active Learning ................................................................. 136
Example 2: Reflective Practice .......................................................... 137
Evaluation of Outcomes ........................................................................ 138
Evaluation of Leadership-Related Competencies ................................. 147
Critical Reflection ................................................................................ 149
References ............................................................................................ 152

Clinical Practice Report 5: Neuropsychological Assessment of Mild Traumatic Brain Injury (TBI): A Case Study of Manoj

Abstract ............................................................................................... 161
References ............................................................................................. 162
List of Appendices

Clinical Practice Report 1
Appendix A. Hospital Anxiety and Depression Scale (HADS, Zigmond & Snaith, 1983)…166
Appendix B. Acceptance of Adjustment to Illness Scale (AIS, Felton et al., 1984)…………168
Appendix C. Work and Social Adjustment Scale (WSAS, Mundt et al., 2002)……………170
Appendix D. Stoma Cognitions Questionnaire (White & Unwin, 1998)………………..171

Clinical Practice Report 2
Appendix E. NRES Guidance: Differentiating Audit, Service Evaluation and Research…..172
Appendix F. Audit Tool…………………………………………………………………………175
Appendix G. NICE CG91 Criteria for ‘Depression in adults with a chronic physical health
problem’ ……………………………………………………………………………………183
Appendix H. Interview Guide……………………………………………………………………193
Appendix I. Clinical Audit Report: X Physical Health Psychology Service and NICE CG91
……………………………………………………………………………………………………195

Clinical Practice Report 3
Appendix J. Anxiety Data……………………………………………………………………214
Appendix K. Supplementary Outcome Data…………………………………………………219
Appendix L. Anxiety Daily Monitoring Sheet………………………………………………220
Appendix M. Anxiety Diary for Specific Episodes…………………………………………221

Clinical Practice Report 4
Appendix N. Training Evaluation Form………………………………………………………222
Clinical Practice Report 5

Appendix O. Neuropsychological Assessment Results........................................223

Appendix P. Case Study Presentation Prezi Slides with Timeline and Formulation.........227
List of Figures

Clinical Practice Report 1

Figure 1. Genogram: Visual representation of Adam’s family tree and relationships.
   Genogram symbols are based on Carr (2006). ....................................................11

Figure 2. Timeline of events in Adam’s life. .................................................................14

Figure 3. Adam's Cognitive Behavioural Formulation using Beck's Longitudinal Model
   (1979)....................................................................................................................16

Figure 4. Adam’s Psychodynamic Formulation using Malan’s Triangles (1979), Triangle of
   Conflict and Triangle of Person.............................................................................21

Clinical Practice Report 2

Figure 1. Visual representation of the structure of the Physical Health Psychology Service, in
   which audited teams are presented in orange.........................................................38

Figure 2. Bar chart displaying age and gender of the audit population.........................47

Figure 3. Pie chart displaying the ethnicity of the audit population.............................48

Figure 4. Pie chart displaying the reason for referral of the audit population...............49

Figure 5. Pie chart displaying the severity of depression of the audit population..........50

Clinical Practice Report 3

Figure 1. Initial formulation based on Carr and McNulty (2006). ..............................90

Figure 2. Formulation based on the Metacognitive Model (Wells, 1995).....................91

Figure 3. Line graph showing daily anxiety frequency and intensity ratings over time....99
Clinical Practice Report 4

Figure 1. Model of Core Implementation Components (Fixsen et al., 2005)………………..127

Figure 2. Presentation slide used for Active Learning element of MI Training. …………137

Figure 3. Presentation slide used for Reflective Practice element of MI Training…………..138

Figure 4. Graphical representation of feedback received on MI Training in relation to Aim 1.
………………………………………………………………………………………………………145

Figure 5. Graphical representation of feedback received on MI Training in relation to Aim 2.
………………………………………………………………………………………………………146

Figure 6. Graph representation of feedback on MI Training for leadership and consultancy
competencies…………………………………………………………………………………148
List of Tables

Clinical Practice Report 1

Table 1. Psychometric Assessment Data...............................................................6

Clinical Practice Report 2

Table 1. Relevant Criteria from NICE CG91 ‘Depression in adults with a chronic physical

health problem’ ........................................................................................................39
Table 2. Collaborative Care (Criterion 14) for Cases meeting 14.1, 14.2, 14.3 and 14.4....51
Table 3. Collaborative Care (Criterion 14) for Cases not meeting 14.1, 14.2, 14.3 and 14.4.53
Table 4. Patient-Centred Care: Patients (Criterion 15).............................................54
Table 5. Patient-Centred Care: Carers (Criterion 16)..............................................55
Table 6. Effective Delivery of Interventions for Depression (Criterion 17)...................56
Table 7. Delivery of Low-Intensity Psychosocial Interventions (Criterion 18)............58
Table 8. Group-Based Peer Support (Self-Help) Programmes (Criterion 25) ..............59
Table 9. Individual Guided Self-Help Programmes (Criterion 20)..............................62
Table 10. Computerised Cognitive Behavioural Therapy (CCBT) (Criterion 21).........63

Clinical Practice Report 3

Table 1. Structure and Content of the Assessment Phase.......................................81
Table 2. Structure and Content of the Intervention Phase........................................93
Table 3. Frequency and Intensity of Anxiety............................................................97
Table 4. Supplementary Outcome Data: HADS and CORE....................................102
Clinical Practice Report 4

Table 1. Characteristics of Motivational Interviewing (Miller & Rollnick, 1991) .............. 116
Table 2. Benefits of CPD (BPS, 2010) ............................................................................. 120
Table 3. Relevant Standards of Proficiency for Practitioner Psychologists (HCPC, 2015) .. 121
Table 4. HCPC Types of CPD Relevant to this Project ....................................................... 125
Table 5. Study Aims ............................................................................................................ 134
Table 6. Quantitative Data from Evaluation Forms ......................................................... 139
Table 7. Qualitative Data from Evaluation Forms ............................................................. 142
VOLUME ONE

Research Component
Literature Review

Theory of Mind Impairment in Individuals who have Experienced Traumatic Brain Injury:

A Meta-Analysis
Abstract

Introduction
Theory of mind (ToM) reasoning is essential for effective social functioning. The literature suggests that deficits in ToM abilities are an underlying cause of impaired social cognition in patients with acquired brain injury (ABI). However, the specific components of ToM that are vulnerable, and the severity of impairments, are not clear. The present meta-analysis reduces risk of bias in the sample by focusing of ABIs with traumatic aetiology.

Method
A systematic literature search of MEDLINE and PsycINFO databases identified 19 eligible articles. The 19 articles consisted of 28 comparisons of TBI and control group performance on ToM tasks: first-order belief (FOTOM) task, second order belief (SOTOM) task, understanding indirect speech (IS) and social faux pas.

Results
The results of the meta-analysis showed medium to large effect sizes, indicating moderate to severe impairment in ToM abilities in individuals who have experienced a TBI. The highest effect size was observed in understanding IS tasks (SMD=0.92), followed by faux pas (SMD=0.83), SOTOM (SMD=0.80) and finally FOTOM (SMD=0.53). Sensitivity analyses revealed that the severity of impairment demonstrated on ToM tasks was influenced by localisation and lateralisation of the injury, whether the injury had traumatic aetiology, time between injury and testing and level of education.

Discussion
The pattern of results indicates partial support for a hierarchical model of ToM development in childhood. Findings demonstrate consistency with frontal lobe localisation of first order belief and second order belief ToM function and right hemispheric lateralisation of more complex faux pas and understanding indirect speech abilities. The clinical implications of findings are discussed in relation to the design and delivery of neuropsychological rehabilitation and individualised care packages.

Conclusions
Important quantitative evidence is presented regarding the severity of, and factors influencing, ToM impairment in individuals who have experienced a TBI.
**Introduction**

Moderate to severe Traumatic Brain Injuries (TBIs) often result in cognitive, emotional and social sequelae (Spikman et al., 2012). Impairments in various components of social cognition have been demonstrated in studies involving neuropsychological assessment of TBI patients, including emotion perception (Bornhofen & McDonald, 2008; Ietswaart, Milders, Crawford, Currie, & Scott, 2008; Milders, Fuchs, & Crawford, 2003) Theory of Mind (ToM) (McDonald and Flanagan, 2004; Milders, Fuchs, & Crawford, 2003) and empathy (Williams & Wood, 2010; Wood & Williams, 2008). These changes post-injury are acknowledged to be among the most devastating of consequences (Spikman et al., 2012). Difficulty in social functioning is more strongly associated with problems in daily life and unfavourable longer-term outcomes than it is with cognitive impairment (Crépeau & Scherzer, 1993; Felmingham, Baguley, & Crooks, 2001; Vilkki et al., 1994). Better understanding of the nature and severity of ToM deficit following TBI has important implications for rehabilitation planning and care management. The present meta-analysis investigates the extent of ToM deficits in individuals with TBI.

**Theory of Mind (ToM).** A core component of social cognition is the ability to infer or reflect on the content of one’s own and other’s mental states. This capacity is known as Theory of Mind (ToM) (Premack & Woodruff, 1978). Inferences of this type allow prediction of other people’s behaviour based on their mental states and are essential for social interactions (Aboulafia-Brakha, Christe, Martory, & Annoni, 2011).

ToM is a multidimensional capacity, involving a range of simple and more complex mentalising skills (Martín-Rodríguez & León-Carrión, 2010; Ubukata et al., 2014). Accordingly, different tools have been designed for its assessment. The majority of ToM assessment methodologies originate from use with children in developmental, educational
and/or clinical psychology settings, but modified tests have been developed for use in other populations, such as adults with an Acquired Brain Injury (ABI).

Belief reasoning tasks are the most widely used assessment of ToM performance in individuals with ABI (Martín-Rodríguez & León-Carrión, 2010). They include first order belief ToM (FOTOM) and second order belief ToM (SOTOM) tasks. FOTOM tasks require the individual to infer someone else’s mental state (e.g. ‘John thinks that object 1 is in location A’) (Wimmer & Perner, 1983). In contrast, SOTOM tasks are more complex and require the individual to identify embedded mental states, demonstrating an understanding that people can have beliefs regarding others’ beliefs (e.g. ‘John thinks that Helen thinks that object 1 is in location A’).

More pragmatic ToM tasks include understanding indirect speech (IS) and the detection of social faux pas. Understanding IS tasks are used to assess an individual’s ability to identify and comprehend irony, sarcasm, metaphors or jokes, in the representation of another individual’s mental state. To complete an understanding IS task an individual is required to reason about mental states in complex social situations. As such, understanding IS tasks are associated with increasing and differential demands on ToM ability in comparison to FOTOM and SOTOM tasks.

In social faux pas tasks the individual is asked to identify whether a character in a story has said something they shouldn’t have, or that is awkward. Tasks of this type include questions that inquire into the emotional state of the character that commits the faux pas. These tasks require more advanced use of ToM inference and sophisticated cognitive processing (McDonald & Flanagan, 2004). Faux pas tasks can be differentiated from FOTOM and
SOTOM by their involvement of emotional information (Shamay-Tsoory, Tomer, Berger, Goldsher, & Aharon-Peretz, 2005; Stone, Baron-Cohen, & Knight, 1998).

As depicted in Figure 1, the development of ToM abilities in childhood shows a fixed and universal sequence, with more complex abilities being developed later (Avis & Harris, 1991; Wimmer & Perner, 1983). It follows that ToM impairment can be explored using a hierarchy of ToM tasks, with each task tapping into different abilities. This hierarchical approach was adopted by ToM research in abnormal psychology: a performance pattern in which an individual passes lower level ToM tasks (such as FOTOM and SOTOM), but fails higher-lever ones (such as understanding IS and social faux pas), has been observed in autistic spectrum disorders and schizophrenia (Corcoran, Mercer, & Frith, 1995; Leslie & Thaiss, 1992). Given the existing literature base in adults with developmental or psychiatric conditions, it is proposed that ToM tasks can be used to systematically assess the nature and severity of ToM deficits in adults with ABI.

Figure 1. The development of ToM abilities in childhood.

As depicted in Figure 1, the development of ToM abilities in childhood shows a fixed and universal sequence, with more complex abilities being developed later (Avis & Harris, 1991; Wimmer & Perner, 1983). It follows that ToM impairment can be explored using a hierarchy of ToM tasks, with each task tapping into different abilities. This hierarchical approach was adopted by ToM research in abnormal psychology: a performance pattern in which an individual passes lower level ToM tasks (such as FOTOM and SOTOM), but fails higher-lever ones (such as understanding IS and social faux pas), has been observed in autistic spectrum disorders and schizophrenia (Corcoran, Mercer, & Frith, 1995; Leslie & Thaiss, 1992). Given the existing literature base in adults with developmental or psychiatric conditions, it is proposed that ToM tasks can be used to systematically assess the nature and severity of ToM deficits in adults with ABI.
Theory of Mind (ToM) and Neuroanatomy. Saxe (2003; 2006) proposed three different brain areas that are critical to ToM: the prefrontal cortex, the posterior cingulate cortex and the bilateral temporo-parietal junction. These three areas demonstrate consistent activation in neuroimaging studies (Fletcher et al., 1995; Ruby & Decety, 2001; Saxe & Kanwisher, 2003), however, their specific contributions to ToM remain unclear.

Impaired performance in ToM tests has been reported in patients with neuroanatomical damage (Bibby & McDonald, 2005; Channon & Crawford, 2000; Channon et al., 2005; McDonald & Flanagan, 2004; Milders, Fuchs, & Crawford, 2003; Milders, Ietswaart, Crawford, & Currie, 2006). When asked to perform mentalising tasks, impaired ToM abilities are demonstrated by patients with highly selective frontal lobe (FL) lesions (Rowe, Bullock, Polkey, & Morris, 2001), anterior capsulotom that interrupt neural pathways between FL and subcortical nuclei (Happé, Malhi, & Checkley, 2001), or selective temporo-parietal junction damage (Samson, Apperly, Chiavarino, & Humphreys, 2004). Such research has contributed to the consistently reported association between FL damage and impairment of social functioning (Rowe, Bullock, Polkey, & Morris, 2001).

Neuropsychological data is inconsistent regarding the role of the left hemisphere (LH) and right hemisphere (RH) in ToM reasoning. Traditionally the RH has been associated with social cognition (Happé, Brownell, & Winner, 1999; Ruby & Decety, 2001), however, neuroimaging studies often show activation in the LH during ToM tasks (Fletcher et al., 1995). Furthermore, in patients with unilateral LH lesions, neuropsychological methodologies have been used to demonstrate ToM deficits in comparison with healthy controls (Channon & Crawford, 2000). Despite this, when LH patient performance was compared with that of RH patients on a faux pas task (e.g. Shamay-Tsoory, Tomer, & Aharon-Peretz, 2005), no
difference was found. Therefore, the current literature base presents conflicting evidence regarding the influence of hemispheric lateralisation on ToM abilities in individuals with TBI.

**Theory of Mind (ToM) and Traumatic Brain Injury (TBI).** The majority of neuropsychological research into ToM has examined an Acquired Brain Injury (ABI)\(^1\) population. Impaired ToM reasoning is widely postulated as an underlying cause of social cognition deficits in patients with ABI (Milders, Fuchs, & Crawford, 2003; Shamay-Tsoory, Tibi-Elhanany, & Aharon-Peretz, 2006; Stone et al., 1998; Stuss, Gallup, & Alexander, 2001). However, the literature does not agree on the specific nature or level of severity of ToM impairment in this clinical population.

Attempting to address this issue, Martín-Rodríguez and León-Carrió (2010) used a meta-analytic approach to review the literature published on ToM performance in ABI between 1995 and 2008. Their review paper included 26 studies which compared the performance of ABI patients with healthy control participants on four widely used ToM tasks: FOTOM, SOTOM, understanding IS and social faux pas. It was reported that overall patients showed moderate to severe ToM impairment; with the highest effect size observed in understanding IS (ES=0.87), followed by faux pas (effect size=0.70), SOTOM (ES=0.60) and finally FOTOM (ES=0.52). The severity of ToM impairment was influenced by the location of injury, type of belief task and heterogeneity of sample aetiology. The relative preservation of FOTOM and SOTOM abilities, compared to more complex faux pas and understanding IS, are in-keeping with a hierarchical model of ToM reasoning in which abilities developed earlier in childhood have relative preservation. However, the presence of extensive

---

\(^1\) ABI is an umbrella term, encompassing TBI and any other non-developmental neural insult (Menon, Schwab, Wright, & Maas, 2010).
overlapping of confidence intervals, and the unexpected finding of a larger effect size for understanding IS than faux pas, means that Martín-Rodríguez and León-Carrión (2010) were unable to provide full support for a hierarchical development model of ToM in ABI.

Martín-Rodríguez and León-Carrión (2010) reported that it is likely that the use of heterogeneous ABI groups in ToM studies (with varying aetiology and severity of conditions) would have dissipated the differences reported in the research, increasing the variability of performance estimates. The pathophysiology of brain damage varies both between and within ABI conditions. TBIs are more likely to involve primary injury characterised by damage resulting from shear forces at impact (Zasler et al., 2012), which may include skull fracture, contusion, subarachnoid or focal haemorrhage, haematoma and/or axonal laceration or shearing (Haydel, 2017). Therefore, ToM studies that use a TBI sample are likely to show a different pattern of cognitive functioning to other types of ABI. Martín-Rodriguez and León-Carrión (2010) included heterogeneity of ABI aetiology in their meta-analysis as a moderator variable, and reported that it had a significant influence on effect sizes for FOTOM and SOTOM tasks. The impact of the proportion of the sample with TBI was explored using weighted regression analysis, which revealed no effect for FOTOM, SOTOM or understanding IS, but a significant trend for faux pas tasks. It seems that the pattern of ToM deficit varies across individuals with TBI and non-traumatic ABI, but the precise pattern of variation is not yet clear.

It was also noted by Martín-Rodríguez and León-Carrión (2010) that that the proportion of patients with FL lesions had a moderating impact on faux pas task performance. Studies with a higher proportion of patients with FL lesions indicated greater differences between patient and healthy control performance on faux pas tasks, but this effect was not observed in other ToM tasks. The analysis also showed significant correlation between the proportion of
patients with RH damage and size of effects for understanding IS and faux pas. Performance on these tasks decreased as the proportion of RH patient variables increased. The existing literature suggests that the localisation and lateralisation of TBI is associated with nature of TBI impairment but more research is needed to clarify the parts of the brain involved in ToM.

**The Need for a Meta-analysis of ToM in TBI Survivors.** No systematic review of the relationship between ABI and ToM impairment has been conducted since 2008. Furthermore, no meta-analysis has been conducted to explore ToM deficits specifically in brain injuries with traumatic aetiology. An updated systematic analysis of the relationship between ABI and ToM impairment, with a specific focus on TBI, is warranted because heterogeneity in injury aetiology may have contributed to the variance in findings in previous research. Furthermore, the inconsistent findings in the existing literature base regarding the influence of TBI hemispheric lateralisation and focal location require elucidation.

A meta-analytic approach was adopted in order to synthesise the available evidence into an objective and structured review of the literature. The meta-analytic approach was chosen as it allows for the estimation of the effect across variations in study methodology and the idiosyncrasies of patient characteristics inherent in the primary studies. Meta-analysis also allows the evaluation of moderator variables on the reported effects (Rosenthal & DiMatteo, 2001). Accordingly, meta-analysis is a useful tool for summarising the existing literature and testing hypotheses regarding the cause of variation between the studies.
Aims

The present meta-analysis aims to extend the current literature, most notably, the previous meta-analysis conducted by Martín-Rodríguez and León-Carrión (2010), by investigating the extent of ToM deficits reported in the literature on TBI.

Following a hierarchical approach to ToM development, it is hypothesized that participants are most likely to demonstrate largest impairment in faux pas, then understanding IS, then SOTOM, and then finally, FOTOM tasks.

The meta-analysis will also examine associated factors that could explain the differences in ToM performance. The following uncontrolled factors, which may impact on effect sizes, will be included as moderator variables: sample characteristics, methodological features and clinical information related to the TBI.

Method

Search Strategy

A systematic literature search was conducted in October 2016 using MEDLINE and PsycINFO databases. As this literature review builds on previous review conducted by Martín-Rodríguez and León-Carrión (2010), the same search terms were used: “theory of mind” [AND] “brain injury”; “theory of mind” [AND] “head injury”; “theory of mind” [AND] “brain damage”. Following on from the work of Martín-Rodriguez and León-Carrión (2010), the search included articles published between June 2008 and October 2016. Only papers published in English were included and the initial search yielded 246 papers. These 246 papers were screened according to the inclusion and exclusion criteria outlined in Table 1.
### Exclusion and Inclusion Criteria

<table>
<thead>
<tr>
<th>Exclusion Criterion</th>
<th>Rationale</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Traumatic Brain Injury (TBI) Group.</strong> The nature of brain injury in the clinical sample group must be Traumatic Brain Injury (TBI) in at least 50% of cases.</td>
<td>To address recommendations aimed at reducing sample heterogeneity (which is likely to dissipate the differences reported in the research, increasing the variability of estimating performance in these individuals).</td>
</tr>
<tr>
<td><strong>Other Theory of Mind (ToM) Tasks.</strong> Studies which used a Theory of Mind (ToM) task other than FOTOM, SOTOM, understanding IS or faux pas were excluded.</td>
<td>This is in accordance with the exclusion criteria established by Martín-Rodriguez and León-Carrión (2010). Character intention tasks were excluded because of their scarce appearance in the literature. In these tasks the participant is required to infer a character’s behaviour based on contextual information. Although this requires FOTOM abilities, the scenario usually provides more detailed information to infer the final character’s behaviour than conventional FOTOM tasks.</td>
</tr>
<tr>
<td><strong>Article Type.</strong> Review articles, meta-analyses, theoretical frameworks or models, clinical guidance papers, commentaries, test development or tool validation papers, single and group case studies, clinical protocols and qualitative studies were excluded.</td>
<td>Articles of these types do not provide the data required for meta-analysis.</td>
</tr>
<tr>
<td><strong>Sufficient Summary Data.</strong> Studies are required to include both means and standard deviations, or sufficient statistical information ($t$, $F$, $X$, $Z$ scores or $p$ values).</td>
<td>Sufficient statistical information was required in order to calculate effect sizes in the clinical and the control groups.</td>
</tr>
</tbody>
</table>
### Inclusion Criterion | Rationale
--- | ---
**Adulthood.** TBI samples were limited to individuals who had experienced their brain injury in adulthood. | This is in accordance with the exclusion criteria established by Martín-Rodríguez and León-Carrión (2010) and reduces heterogeneity across studies. ToM development in children continues up to age 11 and older (Avis & Harris, 1991; Wimmer & Perner, 1983). By studying an adult population it can be assumed that the ToM abilities depicted in Figure 1 will have already been laid down. Therefore, any disruption in these abilities is likely to be associated with subsequent neural trauma.

**Control Group.** Only studies which compared ABI patients’ performance on ToM tasks with a healthy control group were included. | This information is needed to analyse the effect of brain injury on ToM tasks.

**Brain Injury Group.** To be included studies were required to have a clinical sample group of participants who had experienced a brain injury. | This information is needed to analyse the effect of brain injury of ToM tasks.

Following the search strategy, seven eligible articles were identified for inclusion in the meta-analysis. Of the 26 studies included in the analysis conducted by Martín-Rodríguez and León-Carrión (2010), 12 met inclusion criteria for the present study and were carried forward into the analysis, bringing the total number of studies to 19 (see Figure 2).

---

2 Of the residual 14 studies, 12 did not have a sample consisting of greater than 50% TBI, one did not have an adult sample and insufficient data was available for one study.
## Initial Search

| Articles Identified from MEDLINE and PsycINFO Databases | 246 |
| Articles from Martín-Rodríguez and León-Carrión (2010) | 26 |

\[ n=272 \]

## Removed Following Title Screening

<table>
<thead>
<tr>
<th>Category</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Duplicates</td>
<td>70</td>
</tr>
<tr>
<td>Paediatric or Adolescent TBI</td>
<td>32</td>
</tr>
<tr>
<td>Book Chapters or Dissertations</td>
<td>22</td>
</tr>
<tr>
<td>Test Development or Tool Validation</td>
<td>10</td>
</tr>
<tr>
<td>Commentary Article</td>
<td>8</td>
</tr>
<tr>
<td>Single or Group Case Study</td>
<td>7</td>
</tr>
<tr>
<td>Rehabilitation Programme/Clinical Management</td>
<td>6</td>
</tr>
<tr>
<td>Review Article</td>
<td>6</td>
</tr>
<tr>
<td>No ABI Group</td>
<td>4</td>
</tr>
<tr>
<td>No Control Group</td>
<td>2</td>
</tr>
<tr>
<td>&lt;50% TBI</td>
<td>1</td>
</tr>
<tr>
<td>Music Emotion Recognition</td>
<td>1</td>
</tr>
</tbody>
</table>

## Remaining Articles \[ n=103 \]

## Removed Following Abstract Screening

<table>
<thead>
<tr>
<th>Category</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initially Published Before 2008</td>
<td>24</td>
</tr>
<tr>
<td>Other ToM Tests</td>
<td>16</td>
</tr>
<tr>
<td>Review Article</td>
<td>5</td>
</tr>
<tr>
<td>Case Study or Case Series</td>
<td>4</td>
</tr>
<tr>
<td>No Control Group</td>
<td>4</td>
</tr>
<tr>
<td>Framework or Theoretical Model</td>
<td>4</td>
</tr>
<tr>
<td>Not English Language</td>
<td>3</td>
</tr>
<tr>
<td>Paediatric or Adolescent TBI</td>
<td>2</td>
</tr>
<tr>
<td>&lt;50% TBI</td>
<td>2</td>
</tr>
<tr>
<td>Additional Duplicates</td>
<td>2</td>
</tr>
<tr>
<td>Commentary Article</td>
<td>1</td>
</tr>
<tr>
<td>No ABI Group</td>
<td>1</td>
</tr>
</tbody>
</table>

## Remaining Articles \[ n=35 \]

## Removed Following Text Screening

<table>
<thead>
<tr>
<th>Category</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;50% TBI</td>
<td>12</td>
</tr>
<tr>
<td>Basic Data Not Reported</td>
<td>3</td>
</tr>
<tr>
<td>Paediatric or Adolescent TBI</td>
<td>1</td>
</tr>
</tbody>
</table>

## Articles Included in Meta-Analysis \[ n=19 \]

*Figure 2. Flowchart of meta-analysis search strategy and results.*
The sample sizes and the standardised mean differences between TBI and control participants (Cohen’s $d$) of the studies included in the meta-analysis are presented in Table 2. It was common for studies to report outcomes for TBI and control groups on more than one type of ToM task. The literature search for the present meta-analysis identified seven suitable studies, which consisted of 10 brain injury and control group comparisons. These were added to the 18 comparisons across 12 studies from the analysis conducted by Martín-Rodríguez and León-Carrión (2010). This resulted in a total of 28 brain injury and control group comparisons (from 19 studies) being included in the quantitative synthesis (see Table 2).

---

3 Muller, Simion, Reviriego, Galera et al. (2010) was the only study to include all four ToM tasks and as such have contributed four effect sizes to the meta-analysis. Shamay-Tsoory was the main author on eight task manipulations, across five studies. Shamay-Tsoory et al.’s research contributed three effect sizes to FOTOM, two to SOTOM, two to understanding IS and one to the faux pas task data.
### Table 2

**Effect Sizes Reported by the Studies Included in the Meta-Analysis**

<table>
<thead>
<tr>
<th>Study</th>
<th>TBI n</th>
<th>Control n</th>
<th>Cohen's d</th>
<th>Standard Error</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>First Order Belief Task</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Channon, Pellieff, and Rule (2005)</td>
<td>18</td>
<td>19</td>
<td>1.04</td>
<td>0.60</td>
<td>0.04</td>
</tr>
<tr>
<td>Martin and McDonald (2005)</td>
<td>16</td>
<td>16</td>
<td>0.54</td>
<td>0.36</td>
<td>0.13</td>
</tr>
<tr>
<td>McDonald and Flanagan (2004)</td>
<td>34</td>
<td>34</td>
<td>1.15</td>
<td>0.26</td>
<td>0.07</td>
</tr>
<tr>
<td>Muller, Simion, Reviriego, Galera et al. (2010)</td>
<td>15</td>
<td>15</td>
<td>0.74</td>
<td>0.37</td>
<td>0.14</td>
</tr>
<tr>
<td>Shamay-Tsoory and Aharon-Peretz (2007)</td>
<td>49</td>
<td>44</td>
<td>0.36</td>
<td>0.21</td>
<td>0.04</td>
</tr>
<tr>
<td>Shamay-Tsoory, Tibi-Elhanany, and Aharon-Peretz (2006)</td>
<td>44</td>
<td>18</td>
<td>0.21</td>
<td>0.28</td>
<td>0.08</td>
</tr>
<tr>
<td>Shamay-Tsoory, Tibi-Elhanany, and Aharon-Peretz (2007)</td>
<td>48</td>
<td>35</td>
<td>0.17</td>
<td>0.22</td>
<td>0.05</td>
</tr>
<tr>
<td><strong>Second Order Belief Task</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Martin and McDonald (2005)</td>
<td>16</td>
<td>16</td>
<td>1.31</td>
<td>0.39</td>
<td>0.15</td>
</tr>
<tr>
<td>McDonald and Flanagan (2004)</td>
<td>34</td>
<td>34</td>
<td>1.12</td>
<td>0.26</td>
<td>0.07</td>
</tr>
<tr>
<td>Muller, Simion, Reviriego, Galera et al. (2010)</td>
<td>15</td>
<td>15</td>
<td>1.04</td>
<td>0.39</td>
<td>0.15</td>
</tr>
<tr>
<td>Shamay-Tsoory, Tibi-Elhanany, and Aharon-Peretz (2006)</td>
<td>44</td>
<td>18</td>
<td>0.42</td>
<td>0.31</td>
<td>0.09</td>
</tr>
<tr>
<td>Shamay-Tsoory, Tibi-Elhanany, and Aharon-Peretz (2007)</td>
<td>48</td>
<td>35</td>
<td>0.34</td>
<td>0.22</td>
<td>0.05</td>
</tr>
<tr>
<td><strong>Understanding Indirect Speech Task</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bara, Tirassa, and Zettin (1997)</td>
<td>13</td>
<td>13</td>
<td>1.07</td>
<td>0.42</td>
<td>0.18</td>
</tr>
<tr>
<td>Channon, Pellieff, and Rule (2005)</td>
<td>18</td>
<td>19</td>
<td>1.61</td>
<td>0.38</td>
<td>0.14</td>
</tr>
<tr>
<td>McDonald and Pearce (1996)</td>
<td>10</td>
<td>20</td>
<td>1.51</td>
<td>0.44</td>
<td>0.19</td>
</tr>
<tr>
<td>Muller, Simion, Reviriego, Galera et al. (2010)</td>
<td>15</td>
<td>15</td>
<td>1.15</td>
<td>0.39</td>
<td>0.16</td>
</tr>
<tr>
<td>Shamay-Tsoory and Aharon-Peretz (2007)</td>
<td>49</td>
<td>44</td>
<td>0.28</td>
<td>0.21</td>
<td>0.04</td>
</tr>
<tr>
<td>Shamay-Tsoory, Tomer, and Aharon-Peretz (2005)</td>
<td>41</td>
<td>17</td>
<td>0.62</td>
<td>0.09</td>
<td>0.09</td>
</tr>
<tr>
<td><strong>Faux Pas Task</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bivona et al. (2015)</td>
<td>20</td>
<td>20</td>
<td>0.96</td>
<td>0.33</td>
<td>0.11</td>
</tr>
<tr>
<td>Bivona et al. (2014)</td>
<td>28</td>
<td>28</td>
<td>1.56</td>
<td>0.31</td>
<td>0.09</td>
</tr>
<tr>
<td>Geraci Surian, Ferraro, and Cantagallo, (2010)</td>
<td>18</td>
<td>20</td>
<td>1.56</td>
<td>0.37</td>
<td>0.14</td>
</tr>
<tr>
<td>Leopold, Krueger, dal Monte, Pardini et al. (2012)</td>
<td>30</td>
<td>55</td>
<td>0.53</td>
<td>0.22</td>
<td>0.05</td>
</tr>
<tr>
<td>Milders, Fuchs, and Crawford (2003)</td>
<td>17</td>
<td>17</td>
<td>1.06</td>
<td>0.37</td>
<td>0.13</td>
</tr>
<tr>
<td>Milders, Ietswaart, Crawford, and Currie (2006)</td>
<td>30</td>
<td>31</td>
<td>0.32</td>
<td>0.30</td>
<td>0.09</td>
</tr>
<tr>
<td>Muller, Simion, Reviriego, Galera et al. (2010)</td>
<td>15</td>
<td>15</td>
<td>1.16</td>
<td>0.40</td>
<td>0.16</td>
</tr>
<tr>
<td>Shamay-Tsoory, Tomer, Berger, and Aharon-Peretz (2003)</td>
<td>42</td>
<td>19</td>
<td>0.68</td>
<td>0.31</td>
<td>0.10</td>
</tr>
<tr>
<td>Spikman et al. (2012)</td>
<td>28</td>
<td>33</td>
<td>0.60</td>
<td>0.26</td>
<td>0.07</td>
</tr>
<tr>
<td>Yeh, Lo, Tsai, and Tsai (2015)</td>
<td>23</td>
<td>19</td>
<td>0.22</td>
<td>0.32</td>
<td>0.10</td>
</tr>
</tbody>
</table>

Studies which included more than one ToM task are signified using the following system:

\[ ^2 = \text{two tasks contributed by author}; ^3 = \text{three tasks contributed by author}; ^4 = \text{four tasks contributed by author}. \]
**Quality Framework**

In order to identify and assess any risks of bias in the studies, a quality framework was developed and tailored to the specific needs of the present meta-analysis. The quality criteria were adapted from existing frameworks published by Downs and Black (1998), Richards, Jones, Groves, Moss, and Oliver (2015), The Cochrane Collaboration Risk of Bias Tool (Higgins et al., 2011) and the Risk of Bias Assessment Tool for Nonrandomised Studies (RoBANS) (Kim et al., 2013). The quality framework assessed risk of bias in five domains: Selection Bias, Methodological Bias, Measurement Bias, Statistical Bias and Reporting Bias (see Table 3). Each of these five domains consisted of a number of specific items, which were rated as 0, 0.5 or 1 (low, moderate and high risk, respectively) for each study. To reflect their methodological importance, items 11 and 12 were given double weighting, and allocated ratings of 0, 1 or 2 (low, moderate and high risk, respectively). In total, 24 items were rated for each study.
Table 3

Risk of Bias Domains and Items in the Quality Framework

<table>
<thead>
<tr>
<th>Domain</th>
<th>Item</th>
<th>Low, Moderate and High Risks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Selection Bias</td>
<td>1. Representativeness of method of recruitment of ABI group</td>
<td>0 = Random/ multiple sites in multiple cities</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.5 = Multiple sites in 1 city</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1 = 1 site or not reported</td>
</tr>
<tr>
<td></td>
<td>2. Representativeness of method of recruitment of control group</td>
<td>0 = Random/ multiple sites in multiple cities</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.5 = Multiple sites in 1 city</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1 = 1 site or not reported</td>
</tr>
<tr>
<td></td>
<td>3. Demographic variance of participants e.g. all veterans</td>
<td>0 = Heterogeneous/ matched</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.5 = Somewhat homogenous</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1 = Not reported/ homogenous</td>
</tr>
<tr>
<td></td>
<td>4. Matching of gender distribution between control and TBI group</td>
<td>0 = Matched</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.5 = Different but not significantly so</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1 = Not reported/ significantly different</td>
</tr>
<tr>
<td></td>
<td>5. Matching of age distribution between control and TBI group</td>
<td>0 = Matched</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.5 = Different but not significantly so</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1 = Not reported/ significantly different</td>
</tr>
<tr>
<td></td>
<td>6. Matching of education level between control and TBI group</td>
<td>0 = Matched</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.5 = Different but not significantly so</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1 = Not reported/ significantly different</td>
</tr>
<tr>
<td></td>
<td>7. Reporting of clinical characteristics of ABI (e.g. location, open/closed; aetiology)</td>
<td>0 = Aetiology and characteristics reported</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.5 = Aetiology/ characteristics reported</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1 = Not reported</td>
</tr>
<tr>
<td></td>
<td>8. Reporting of co-morbidities (e.g. were individuals with psychiatric conditions excluded?)</td>
<td>0 = Several groups excluded</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.5 = Depression/ mental health/ neurological conditions/excluded</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1 = Not reported/ sig difs between groups</td>
</tr>
<tr>
<td></td>
<td>9. Reporting of time (since injury) of testing</td>
<td>0 = Range and mean reported</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.5 = Mean/ minimum criteria/ range reported</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1 = Not reported</td>
</tr>
<tr>
<td></td>
<td>10. Difference in range of</td>
<td>0 = &lt;10 years</td>
</tr>
<tr>
<td>Methodological Bias</td>
<td>Time (since injury) of testing</td>
<td>0.5 = 10 to 20 years</td>
</tr>
<tr>
<td>---------------------</td>
<td>--------------------------------</td>
<td>---------------------</td>
</tr>
<tr>
<td></td>
<td>1 = Not reported/ &gt;20 years</td>
<td></td>
</tr>
<tr>
<td>11. Proportion of ABI group that experienced a TBI*</td>
<td>0 = 84-100% TBI</td>
<td>1 = 67-83% TBI</td>
</tr>
<tr>
<td></td>
<td>2 = 50-66% TBI/ not reported</td>
<td></td>
</tr>
<tr>
<td>12. Sample size (total n)*</td>
<td>0 = &gt;60</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1 = 30 to 60</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2 = Not reported/ &lt;30</td>
<td></td>
</tr>
<tr>
<td>13. Matching of sample size (difference between size of ABI group and control group)</td>
<td>0 = -10 to 10</td>
<td>0.5 = -20 to -10/ 10 to 20</td>
</tr>
<tr>
<td></td>
<td>1 = Not reported/ &gt;20/ &lt;-20</td>
<td></td>
</tr>
<tr>
<td>14. Was there standard administration or any changes to administration (e.g. shortened batteries)? Based on previous research?</td>
<td>0 = Standardised administration</td>
<td>0.5 = Minor changes</td>
</tr>
<tr>
<td></td>
<td>1 = Not reported</td>
<td></td>
</tr>
<tr>
<td>15. Researcher blinding</td>
<td>0 = Full blinding</td>
<td>0.5 = Some blinding (i.e. some researchers)</td>
</tr>
<tr>
<td></td>
<td>1 = Not reported/ no blinding</td>
<td></td>
</tr>
<tr>
<td>16. Were the tests delivered in groups or 1-1?</td>
<td>0 = 1:1</td>
<td>0.5 = Mixture of group and 1:1</td>
</tr>
<tr>
<td></td>
<td>1 = Not reported</td>
<td></td>
</tr>
<tr>
<td>17. Reporting of the test-retest reliability</td>
<td>0 = Good reliability</td>
<td>0.5 = Adequate reliability</td>
</tr>
<tr>
<td></td>
<td>1 = Not reported/ poor reliability</td>
<td></td>
</tr>
<tr>
<td>18. Reporting of the reliability of the ToM Measure</td>
<td>0 = Good reliability</td>
<td>0.5 = Adequate reliability</td>
</tr>
<tr>
<td></td>
<td>1 = Not reported/ poor reliability</td>
<td></td>
</tr>
<tr>
<td>19. Reliability and standardisation of scoring of the ToM task</td>
<td>0 = Standardised scoring</td>
<td>0.5 = Minor changes</td>
</tr>
<tr>
<td></td>
<td>1 = Not reported/ major changes</td>
<td></td>
</tr>
</tbody>
</table>
20. Have counterbalancing/order effects/fatigue been addressed?

- 0 = Addressed
- 0.5 = Partially addressed
- 1 = Not reported/ not addressed

21. Missing or incomplete data (e.g. the n in one section is different to the n in another section of the report)

- 0 = No important data missing
- 0.5 = Some important data missing
- 1 = Not reported/ important data missing

22. Reporting subgroup data only (e.g. only Right Hemisphere were significant)

- 0 = No important data missing
- 0.5 = Some important data missing
- 1 = Not reported/ important data missing

23. Not being able to find important information in the paper

- 0 = Able to find all info
- 0.5 = Able to find most info
- 1 = Able to find some or no info

24. Inconsistencies in reporting of the data

- 0 = No important data missing
- 0.5 = Some important data missing
- 1 = Not reported/ important data missing

*Factors allocated double weighting

**Quality Ratings**

From the individual risk ratings, the mean was calculated for each study, to provide an overall index of risk of bias. These overall ratings for each study are presented in Table 4, with lower ratings indicating less risk of bias and a more robust study. The overall ratings ranged from 0.25 to 0.63 with the lowest quality study conducted by Shamay-Tsoory, Tomer, and Aharon-Peretz (2005) with an understanding IS task and the highest quality study conducted by Milders et al. (2006) with a faux pas task.

Overall assessment of the primary studies indicated that the Measurement Bias domain showed the highest proportion of ‘high risk’ ratings; a high number of studies failed to report the psychometric properties of the ToM tasks implemented. In contrast, the most common rating for items of Selection Bias was ‘low risk’, with many studies selecting a relatively large and representative sample (see Table 4).
### Table 4

**Risk of Bias Ratings Based on Quality Framework**

<table>
<thead>
<tr>
<th>Study</th>
<th>Selection Bias (1-11)</th>
<th>Methodological Bias (12-16)</th>
<th>Measurement Bias (17-20)</th>
<th>Statistical Bias (21)</th>
<th>Reporting Bias (22-24)</th>
<th>Overall Quality Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>First Order Tasks</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Channon et al. (2005)</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>Shamay-Tsoory and Aharon-Peretz (2007)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shamay-Tsoory et al. (2007)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shamay-Tsoory et al. (2006)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Martin and McDonald (2005)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>McDonald and Flanagan (2004)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Muller et al. (2010)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Second Order Tasks</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shamay-Tsoory et al. (2007)</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>Shamay-Tsoory et al. (2006)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Martin and McDonald (2005)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>McDonald and Flanagan (2004)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Muller et al. (2010)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Understanding IS Tasks</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>-------------------------------------</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
</tr>
<tr>
<td>Bara et al. (1997)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Channon et al. (2005)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shamay-Tsoory and Aharon-Peretz (2007)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>McDonald and Pearce (1996)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shamay-Tsoory et al. (2005)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Muller et al. (2010)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Faux Pas Tasks</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Milders et al. (2006)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shamay-Tsoory et al. (2003)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Milders et al. (2003)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Geraci et al. (2010)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Muller et al. (2010)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yeh et al. (2015)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leopold et al. (2012)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spikman et al. (2012)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bivona et al. (2015)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bivona et al. (2014)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

= High Risk  = Moderate Risk  = Low Risk
**Procedure for Quantitative Synthesis**

Separate meta-analyses were conducted for each ToM task, using the “Meta” package in R (Schwarzer, 2007; Schwarzer, Carpenter, & Rucker, 2015; R Core Team, 2015), with meta-analyses a function of methodological and idiosyncratic factors.

A random effects model was used to calculate the meta-analytic effect for each ToM test. Random effects models assume that effect sizes across studies may vary; as such it aims to estimate the mean and distribution of effects. In contrast, fixed effects models assume the true effect size for all studies is identical, although the measured effect size may vary between studies due to sampling error. It is assumed that a random effects model is more appropriate than a fixed effects model when researching psychological variables because the measured effects in psychological research are likely to vary as a result of a number of uncontrolled factors\(^4\). It is likely that the psychological studies reported in this meta-analysis will produce differing estimates of the relationship between TBI and ToM deficit as a function of participant idiosyncrasies, methodological variations and discrepancies in test administration.

In addition to the random effects model, a quality effects model was calculated. The random effects model assumes that precision is a function of sample size and weights studies accordingly, with larger samples receiving a greater weighting in the analysis. In contrast, the quality effects model acknowledges that risk of bias may proceed from multiple sources and that studies of greater methodological quality (and lower risk of bias) produce more precise estimates of effect size. Accordingly, the quality effects model weights against studies with a

\(^4\) Psychological studies are often carried out by independent researchers who implement different research methodologies, use varying definitions of psychological concepts and implement a range of tests (with differing psychometric properties).
greater risk of bias (as measured by the overall quality index score), allowing for estimation of the attenuation of effect due to sub-optimal methodology.

**Calculating Summary Effects.** The meta-analysis could not be performed directly on the raw difference in means, due to different studies in the analysis using different variants of the scales and/or administration procedures. In order to create an index that would be comparable across studies, the mean difference between the TBI and control group in each study was divided by the pooled standard deviation for that study, producing a standardised mean difference (SMD). The sample estimate of the SMD is often referred to as Cohen’s $d$ in research synthesis. To calculate Cohen’s $d$, the mean, standard deviation and number of participants for each study are required.

**Calculating Trial Level Effects.** The values in the tables below are expressed as Cohen’s $d$. Cohen’s $d$ has been shown to systematically overestimate the absolute value of the SMD in small samples (Borenstein, 2009). This bias can be removed by using an unbiased estimate known as Hedge’s $g$ (Hedges, 1981). Therefore, a correction factor was applied, and Hedge’s $g$ values were used for the calculations in the present analysis. Values were transformed back into Cohen’s $d$ for interpretation, and are reported as Cohen’s $d$ in the tables of this meta-analysis.

**Calculating Trial Level Confidence Intervals.** The precision of the calculated effect is represented by the 95% confidence interval (CI). Under a fixed effect model the only source of bias is sampling or estimation error within the study. Under a random effects model, the

---

5 Not all studies in the present meta-analysis reported the basic statistics required to calculate Cohen’s $d$. In such cases Cohen’s $d$ was calculated using different computational strategies. When subgroup means and standard deviations were published, these were combined. If t or f values, and exact $p$ values were reported, these values were converted into Cohen’s $d$. For each ToM task a sensitivity analysis was conducted to assess the impact of using a different computational method for calculating Cohen’s $d$. 


same sampling or estimation error exists, in addition to other potential sources of variation between studies. Accordingly, the CIs reported using the random effects model are often larger than those calculated using the fixed effects model.

**Quantifying Heterogeneity**

Heterogeneity in meta-analyses refers to the diversity in design and conduct of studies, which may or may not be responsible for observed discrepancies in outcomes across studies (Higgins & Thompson, 2002). In order to quantify heterogeneity in the reported analysis, Cochrane’s Q, Higgins $I^2$ and Tau were computed across each of the ToM tasks using the inverse-variance method. High levels of heterogeneity suggest that uncontrolled factors may be influencing the relationship between TBI and ToM, and if found, the possible sources for this heterogeneity should be examined.

Borenstein et al. (2009) recommends the use of Higgins $I^2$ to decide whether or not further exploration of study heterogeneity is required. Higgins, Thompson, Deeks, and Altman (2003) provided benchmark values of <50%, <75% and >75% to reflect low, moderate and high heterogeneity respectively. Given the relatively low levels of experimental control and high levels of participant heterogeneity in clinical research it was decided that a Higgins $I^2$ of 75% or higher would be considered as unacceptable heterogeneity and subgroup analysis or meta-regression would be conducted to identify the causes of the uncontrolled variance.

**Identifying Influential Studies.** To examine whether any particular studies were exerting a disproportionately high influence on the overall meta-analytic effect, a “one left out” procedure was conducted. By observing the impact of removing each study in turn, this procedure identifies individual studies with a disproportionate influence on the quantitative
synthesis. If omitting a study results in an effect that lies outside of the 95% CI for the complete meta-analysis then that study is deemed to have a disproportionate influence.

**Publication Bias.** Publication bias\(^6\) is a problem for all forms of systematic review; however, meta-analysis provides ways to estimate its likelihood and potential impact. In the present analysis, publication bias was assessed by visual examination of Funnel Plots\(^7\). However, as visual interpretation of Funnel Plots is highly subjective (Terrin et al., 2005), a supplementary statistical test of Funnel Plot asymmetry was conducted (Egger, Smith, Schneider, & Minder, 1997). If publication bias is identified, its impact will be estimated using a Trim and Fill analysis\(^8\) (Duval & Tweedie, 2000) and Fail-safe N\(^9\) will be calculated using the Rosenthal method (Rosenthal, 1979).

**Results**

Significant positive associations were observed across each type of ToM task, with individuals in the clinical group demonstrating poorer performance than healthy controls.

Table 5 presents the results of the meta-analysis for each ToM task type.

---

\(^6\) Publication bias refers to systematic bias in the literature resulting from favouring the publication of statistically significant results. Publication bias can fundamentally skew findings, producing invalid overestimates of the effect.

\(^7\) Funnel Plots are a mechanism for displaying the relationship between precision of estimate (study size) and effect size.

\(^8\) Trim and Fill analysis uses an iterative procedure to obtain an unbiased estimate of effect size. The procedure removes the most extreme small studies from the positive side of the Funnel Plot (trim), re-computing the effect size at each iteration, until the Funnel Plot is symmetric around the new mean effect size. The original studies are then added back into the analysis, with a mirror image for each (fill) (Duval & Tweedie, 2000).

\(^9\) Fail-safe N is the number of null studies required to reduce the observed effect to non-significance.
Table 5

*Meta-Analysis Results Summary by Theory of Mind Task*

<table>
<thead>
<tr>
<th>ToM Task</th>
<th>Number of Studies</th>
<th>Number of Participants</th>
<th>Random Effects Model</th>
<th>Heterogeneity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>ABI Control</td>
<td>Summary Effect</td>
<td>Sig ($p$)</td>
</tr>
<tr>
<td>First Order Belief</td>
<td>7</td>
<td>224 181</td>
<td>0.53</td>
<td>0.0005</td>
</tr>
<tr>
<td>Second Order Belief</td>
<td>5</td>
<td>157 118</td>
<td>0.80</td>
<td>0.0001</td>
</tr>
<tr>
<td>Understanding Indirect Speech</td>
<td>6</td>
<td>146 128</td>
<td>0.92</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Faux Pas</td>
<td>10</td>
<td>251 257</td>
<td>0.83</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>
First Order Belief Tasks

A total of seven studies, of 405 participants, reported associations between TBI and performance on FOTOM tasks. As seen in Figure 3, the random effects model yielded a significant effect (SMD=0.53, 95% CI [0.23;0.83]). This suggests there is a positive relationship between presence of a TBI and impairment on FOTOM tasks. When weighting studies by methodological quality, the effect showed only a 4% increase (SMD=0.55, 95% CI [0.22;0.88]), suggesting that this effect was not adversely influenced by methodological variation in the primary studies.

Low heterogeneity was observed across studies administering the FOTOM (Cochrane’s Q=10.91, p=.091; Tau²=0.0686, Higgins I²=45%) and as such, no subgroup analyses were considered necessary.

![Figure 3. Forest plot of first order belief task studies.](image)

Identifying Influential Studies. Following the “one left out” procedure, omitting each study of SOTOM in turn did not demonstrate that any individual study had a disproportionate influence on the meta-analysis (see Figure 4). The study with the greatest effect was McDonald and Flanagan (2004), removal of which resulted in a reduction of the meta-analytic effect to SMD=0.36, however the 95% CI overlapped the effect size for the complete
analysis. Therefore, the study conducted by McDonald and Flanagan (2004) was not considered disproportionately influential.

![Forest plot of influence of first order belief task studies on meta-analytic effect.](image)

**Figure 4.** Forest plot of influence of first order belief task studies on meta-analytic effect.

**Publication Bias.** In the absence of publication bias, one would expect a Funnel Plot to show studies distributed symmetrically around the mean effect size, given that sampling error is random. The absence of studies in the lower left section of the plot indicates that smaller studies yielding small effects may be missing from the meta-analysis (see Figure 5a).

However due to the small number of studies ($n=7$), the scope for examining the possible influence of publication bias is limited.

![Funnel plot (a) and trim and fill analysis (b) for first order belief task studies.](image)

**Figure 5.** Funnel plot (a) and trim and fill analysis (b) for first order belief task studies.
Because the interpretation of a Funnel Plot is largely subjective, Egger’s Test of asymmetry was conducted, which did not produce a significant result ($p=0.312$). However, it is proposed that tests for Funnel Plot asymmetry should only be conducted if there are enough studies to ensure sufficient power; 10 studies is generally seen as adequate. Given that this analysis includes just 7 studies, the absence of a significant regression cannot be confidently taken as evidence of symmetry.

The results of the Trim and Fill analysis can be seen in Figure 5b, in which 3 unfilled circles represent 3 imputed studies. The uncorrected estimate of effect size was 0.53 (95% CI [0.23;0.83]), whereas the imputed estimate of effect size is slightly smaller than the original analysis (SMD=0.30, 95% CI [-0.05;0.65]). With this adjustment for possible publication bias, there is a slightly reduced association estimated between TBI and performance on FOTOM tasks.

To further address the issue of publication bias, Rosenthal’s method was used to compute the Fail-safe N (Rosenthal, 1979). The Fail-safe N was estimated to be 17, signifying that 17 studies would need to be added to the analysis for the meta-analytic effect to become non-significant. Because the Fail-safe N is relatively small, there is a risk that publication bias may have inflated the observed effect between TBI and performance on FOTOM tasks.

**Summary of FOTOM Task Results.** Using the information gathered regarding the inclusion of null studies, it is estimated that the effect would fall between 0.3 and 0.53. The best estimate from the available data is that there is a medium sized effect of 0.53 between TBI and performance on FOTOM tasks. However, due to the small number of studies included in the analysis, caution is warranted in interpretation of results.
**Second Order Belief Tasks**

A total of five studies, of 275 participants, reported associations between TBI and performance on a SOTOM task. The random effects model yielded a significant effect (SMD=0.80, 95% CI [0.39;1.20]). This suggests there is a positive relationship between presence of a TBI and SOTOM impairment. The quality effects model showed a slight decrease in effect (SMD=0.76, 95% CI [0.35;1.17]), suggesting that the effect was robust to variations in methodological quality. Following the benchmark values proposed by Higgins et al. (2003), the level of heterogeneity would be classified as moderate (Cochrane’s Q=9.14, \( p=0.057 \); \( \tau^2=0.1165 \), Higgins \( I^2=56.3\% \)) and accordingly subgroup analysis was not considered necessary.

![Figure 6](image-url)  Forest plot of second order belief task studies.

**Identifying Influential Studies.** As seen in Figure 7, no individual study was found to have a disproportionate influence on the observed meta-analytic effect. The study with the greatest effect was Shamay-Tsoory, Tibi-Elhanany and Aharon-Peretz (2007) which resulted in a continuation of the meta-analytic effect to SMD=0.95. However, the 95% CI contained the effect size for the overall analysis, therefore this study was not considered disproportionately influential.
Publication Bias. The absence of studies in the lower left section of Figure 8a indicates that smaller studies reporting smaller effects may be missing from the analysis. However due to the small number of studies (n=5), the capacity for examining the influence of publication bias is limited.

In order to address the subjectivity of Funnel Plot interpretation, supplementary analysis using the Egger Test was conducted. The Egger Test did not show significant bias (p=0.136), and therefore did not provide evidence to support the presence of publication bias. Given the small number of studies present, and the associated possibility of low power, caution is warranted in interpretation of this analysis.
Figure 8b presents the results of the Trim and Fill analysis. It can be seen that two unfilled circles have been added, to represent two imputed studies. The uncorrected estimate of effect size was 0.80 (95% CI [0.39;1.20]), whereas the imputed estimate of effect size is slightly smaller than the original analysis (SMD=0.52, 95% CI [0.08;0.97]). Accordingly, the adjusted estimate suggests a slightly lower association between TBI and performance on SOTOM tasks.

To further explore the possibility of publication bias, Rosenthal’s method was used to compute the Fail-safe N (Rosenthal, 1979). The Fail-safe N was estimated to be two, signifying that the addition of just two studies would reduce the meta-analytic effect to non-significance. The very small Fail-safe N is indicative of a risk of publication bias. However, these results should be interpreted with caution and any conclusions should be drawn with consideration of the small number of studies in this analysis.

**Summary of SOTOM Task Results.** From the data that is available, it is estimated that there is a large effect size between TBI and SOTOM task performance of 0.80. However, due to the small number of studies included in the analysis of SOTOM studies, and the suggestion of a risk of publication bias, the extent to which conclusions can be drawn is limited. In order to confirm whether this effect is robust, more data is required.
Understanding Indirect Speech Tasks

A total of six studies, of 274 participants, reported associations between TBI and performance on understanding IS tests of ToM. The random effects model yielded a significant effect (SMD=0.92, 95% CI [0.53;1.31]). This suggests there is a positive relationship between presence of a TBI and impairment understanding IS. When weighting studies by methodological quality, the effect increased by 12% (SMD=1.03, 95% CI [0.62;1.43]). This suggests that the effect was mildly influenced by methodological variation in the studies but the difference between the random effects model and the quality effects model was not statistically significant.

The level of heterogeneity would be considered as moderate (Cochrane’s Q=16.06, p=.0067; Tau²=0.1412, Higgins I²=68.9%) and accordingly subgroup analyses were not conducted.

![Figure 9. Forest plot of understanding indirect speech studies.](image)

Identifying Influential Studies. As shown in Figure 10, leaving out each study in turn indicated that no particular study had a disproportionate influence on the overall observed meta-analytic effect. The study with the greatest effect was Shamay-Tsoory and Aharon-Peretz (2007) which resulted in a continuation of the meta-analytic effect to SMD=1.11. However the 95% CI contained the effect size for the overall analysis, therefore this study was not considered disproportionately influential.
Figure 10. Forest plot of influence of understanding indirect speech task studies on meta-analytic effect.

**Publication Bias.** The subjective impression is that there is a lack of studies in the lower left section of the plot (see Figure 11a). This suggests that smaller studies reporting smaller effects may be missing from the analysis, but the small number of studies (n=6), limits the extent to which publication bias can be examined.

![Funnel plot](image)

**Figure 11.** Funnel plot (a) and trim and fill analysis (b) for understanding indirect speech studies.

Egger’s test of Funnel Plot asymmetry did not yield a significant result ($p=0.136$). Caution is warranted in interpretation of the Funnel Plot asymmetry test. If only a small number of studies are present it is possible that the power of the tests will be too low to distinguish chance from real asymmetry.
The results of the Trim and Fill analysis can be seen in Figure 11b, in which three unfilled circles represent three imputed studies. The uncorrected estimate of effect size was 0.92 (CI 95%, [0.53;1.31]). Although remaining statistically significant, the adjusted estimate of effect size is considerably smaller than the original analysis (SMD=0.60, 95% CI [0.21;0.97]).

Using Rosenthal’s method, the Fail-safe N was estimated to be 33. Accordingly, 33 studies would be required for the cumulative effect to become statistically non-significant. Given extensive nature of the literature search, it is considered unlikely that 33 unpublished non-significant studies were missed. Therefore, the assessment of this effect size had not been adversely affected by publication bias, however, the true effect size is likely to be statistically significant and fall between 0.60 and 0.92.

**Summary of Understanding IS Task Results.** The results indicate that if the methodological quality of studies was improved and unpublished null data was made available, the true association between TBI and understanding IS task performance would fall between 0.60 and 0.92. From the available data it is estimated that there is a large effect between TBI and understanding IS task performance, of 0.92.
Faux Pas Tasks

A total of 10 studies, of 508 participants, reported associations between TBI and performance on a faux pas test of ToM. The random effects model yielded a significant effect (SMD=0.83, 95% CI [0.54;1.12]). This suggests there is a positive relationship between presence of TBI and impairment on faux pas tasks. The quality effects model showed a similar effect (SMD=0.89, 95% CI [0.59;1.19]) suggesting that the effect was robust to variations in methodological quality.

Following the benchmark values proposed by Higgins et al (2003), the level of heterogeneity would be classified as moderate and accordingly subgroup analyses were not necessary (Cochrane’s Q=20.04, p=.0177; Tau²=0.1169, Higgins I²=55.1%).

![Figure 12](image.png)

Figure 12. Forest plot of faux pas task studies.

Identifying Influential Studies. As shown in Figure 13, omitting any individual study did not cause the meta-analytic effect to fall outside of the 95% CIs. The study with the greatest effect was Bivona et al. (2014), which resulted in a meta-analytic effect of 0.73. Due to the 95% CI containing the effect size for the overall analysis, this study was not considered.
disproportionately influential. As such, it can be concluded that no one single study had a
disproportionate influence on the meta-analytic effect.

<table>
<thead>
<tr>
<th>Study</th>
<th>Standardised mean difference</th>
<th>SMD</th>
<th>95%-CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Omitting Bivona et al. (2015)</td>
<td>0.82</td>
<td>[0.50; 1.14]</td>
<td></td>
</tr>
<tr>
<td>Omitting Bivona et al. (2014)</td>
<td>0.73</td>
<td>[0.47; 0.99]</td>
<td></td>
</tr>
<tr>
<td>Omitting Geraci, Surian, Ferraro and Cantagalio 2010</td>
<td>0.76</td>
<td>[0.48; 1.03]</td>
<td></td>
</tr>
<tr>
<td>Omitting Leopold, Krueger, dal Monte, Pardini et al 2012</td>
<td>0.87</td>
<td>[0.65; 1.20]</td>
<td></td>
</tr>
<tr>
<td>Omitting Milders et al. (2003)</td>
<td>0.81</td>
<td>[0.50; 1.12]</td>
<td></td>
</tr>
<tr>
<td>Omitting Milders et al. (2006)</td>
<td>0.89</td>
<td>[0.59; 1.19]</td>
<td></td>
</tr>
<tr>
<td>Omitting Muller, Simon, Reviniego, Galera et al 2010</td>
<td>0.80</td>
<td>[0.49; 1.11]</td>
<td></td>
</tr>
<tr>
<td>Omitting Shamay-Tsoory, Tomer, Berger &amp; Aharon-Peretz (2003)</td>
<td>0.85</td>
<td>[0.53; 1.17]</td>
<td></td>
</tr>
<tr>
<td>Omitting Spikman et al. (2012)</td>
<td>0.86</td>
<td>[0.54; 1.19]</td>
<td></td>
</tr>
<tr>
<td>Omitting Yeh, Lo, Tsai and Tsai 2015</td>
<td>0.89</td>
<td>[0.60; 1.19]</td>
<td></td>
</tr>
</tbody>
</table>

Random effects model  

\[
0.83 \ [0.54; 1.12]
\]

Figure 13. Forest plot of influence of faux pas task studies on meta-analytic effect.

**Publication Bias.** The absence of studies in the lower left part of the plot is suggestive that
smaller studies reporting smaller effects could be missing from the meta-analysis (see Figure
14a).

Figure 14. Funnel plot (a) and trim and fill analysis (b) for faux pas task studies.
Given the subjective nature of relying on the author’s impression of a Funnel Plot, Egger’s test for small study effects was conducted (Egger et al., 1997). This linear regression test of Funnel Plot asymmetry did not reveal significant publication bias ($p=0.115$). However, due to the small number of studies present, and the associated possibility of low power, the extent to which conclusions can be drawn from this analysis are limited.

The results of the Trim and Fill analysis can be seen in Figure 14b, in which observed studies are shown as filled circles and imputed studies are shown as unfilled circles. The uncorrected estimate of effect size was 0.83 (95% CI [0.54;1.12]), whereas the imputed estimate of effect size was slightly smaller than the original analysis (SMD=0.60, 95% CI [0.27;0.93]).

Using Rosenthal’s method, the Fail-safe N was estimated to be 84, indicating that 84 studies would need to be added to the analysis for the meta-analytic effect to become statistically non-significant. Given that the literature search was relatively extensive, it is improbable that 84 unpublished non-significant studies were missed. Therefore, it is unlikely that the effect is significantly overestimated due to publication bias.

**Summary of Faux Pas Task Results.** Overall, with improved methodological quality of studies and unpublished null data made available, the true association between TBI and faux pas task performance would fall between 0.6 and 0.83. Accordingly, the best current estimate is that there is a large effect between TBI and faux pas task performance, of 0.83.

**Author Bias**

Shamay-Tsoory’s research contributed three effect sizes to FOTOM, two to SOTOM, two to understanding IS and one to the faux pas task data. Accordingly, there is a possibility that a bias from Shamay-Tsoory’s research could have an effect on the results of the present meta-
analysis. In order to address this issue, random effect models were repeated with studies authored by Shamay-Tsoory excluded. This analysis resulted in an effect size of 0.85 (95% CI [0.53;1.17]) for faux pas tasks, 0.90 (95% CI [0.55;1.25]) for FOTOM, 1.15 (95% CI [0.77;1.52]) for SOTOM and 1.34 (95% CI [0.94;1.74]) for understanding IS tasks. Given the overlapping of CIs, this data does not support the existence of an author bias of the work of Shamay-Tsoory.

*Sensitivity Analyses*

In order to investigate the possible impact of methodological factors and sample characteristics on the meta-analytic effect, sensitivity analyses were conducted. To minimise the risk of bias and so that direct comparisons were more applicable, the variables identified for sensitivity analysis were based on those included in the analysis conducted by Martín-Rodríguez and León-Carrión (2010). Following recommendations made by Martín-Rodríguez and León-Carrión (2010), the variables of proportion of sample that had a TBI, and time between injury and testing (mean and range), were also included in the analysis. For the continuous variables (age, years in education, proportion of males, proportion TBI, proportion RH injury, proportion FL injury, mean time since injury and range in time since injury) meta-regression analyses were conducted. The results of these analyses are presented in Table 6, alongside the results of Martín-Rodríguez and León-Carrión (2010). For the categorical variable of TBI severity, subgroup analysis was conducted and is presented in Table 7\(^{10}\).

\(^{10}\) Results from Martín-Rodríguez and León-Carrión (2010) for severity of condition could not be presented in Table 7 as this variable was not controlled for in their meta-analysis.
Demographic Moderator Variables

Meta-regression analysis did not reveal a significant effect of age on performance in any of the tasks. Similarly, the proportion of males in the sample did not have a significant effect on ToM task performance for FOTOM, understanding IS or faux pas tasks; there was insufficient data available to conduct meta-regression analysis on the impact of age on SOTOM task performance.

The analysis of the impact of years in education on ToM task performance demonstrated a significant effect for faux pas tasks ($p<0.05$), but this effect was not observed for FOTOM or understanding IS tasks. There was insufficient data available to conduct a meta-regression analysis investigating the impact of years in education on SOTOM tasks.

Clinical Moderator Variables

One of the inclusion criteria for the meta-analysis was that at least 50% of cases in the clinical sample had experienced a brain injury that was traumatic in aetiology. Meta-regression analysis of proportion of the clinical sample with a TBI revealed a significant effect for FOTOM ($p<0.01$), SOTOM ($p<0.01$) and understanding IS ($p<0.001$) tasks, however this effect was not observed in faux pas tasks ($p=0.74$).

Meta-regression showed a significant effect of proportion of the clinical sample with FL injury on FOTOM ($p<0.05$) and SOTOM ($p<0.05$) tasks, but this effect was not observed in understanding IS ($p=0.72$) or faux pas ($p=0.39$) tasks.

Meta-regression analysis of lateralisation of TBI indicated a larger effect of TBI on faux pas ($p<0.01$) and understanding IS ($p<0.05$) performance if injuries were RH located. This effect was not observed in FOTOM or SOTOM tasks.
Regarding the mean length of time between date of injury and testing, the meta-regression analysis revealed a significant effect of the length of time since injury on task performance for understanding IS ($p<0.01$) and a trend towards significance for faux pas tasks ($p=0.07$). This suggests that the longer the time since the TBI, the larger the impact on ability for more complex ToM tasks. In terms of the range between the shortest and longest amount of time between injury and testing in the TBI sample, no significant effects were indicated.
Table 6

Results from Meta-regression Analyses of the Impact of Continuous Moderator Variables on ToM Task Effect Sizes for the Present Study (2017) and Previous Meta-analysis by Martin-Rodriguez and León-Carrión (2010)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>-0.1</td>
<td>n.s.</td>
<td>0.06</td>
<td>0.47</td>
<td>-0.15</td>
<td>n.s.</td>
<td>-0.06</td>
<td>n.s.</td>
<td>0.01</td>
<td>0.88</td>
<td>0.54</td>
<td>n.s.</td>
<td>-0.09</td>
<td>0.25</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Education (years)</td>
<td>0.43</td>
<td>n.s.</td>
<td>-0.02</td>
<td>0.93</td>
<td>0.38</td>
<td>n.s.</td>
<td>-0.42</td>
<td>n.s.</td>
<td>-0.14</td>
<td>0.64</td>
<td>-0.32</td>
<td>n.s.</td>
<td>-0.35</td>
<td>0.03*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proportion Male</td>
<td>-0.01</td>
<td>n.s.</td>
<td>0.00</td>
<td>0.93</td>
<td>0.13</td>
<td>n.s.</td>
<td>0.00</td>
<td>0.97</td>
<td>-1.17</td>
<td>n.s.</td>
<td>0.01</td>
<td>0.70</td>
<td>-0.02</td>
<td>0.93</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proportion TBI</td>
<td>-0.05</td>
<td>n.s.</td>
<td>1.85</td>
<td>0.003**</td>
<td>-0.03</td>
<td>n.s.</td>
<td>2.06</td>
<td>0.003**</td>
<td>0.24</td>
<td>n.s.</td>
<td>3.02</td>
<td>0.0003**</td>
<td>0.79</td>
<td>0.06.</td>
<td>0.72</td>
<td>0.74</td>
</tr>
<tr>
<td>Proportion RH</td>
<td>-0.42</td>
<td>n.s.</td>
<td>-0.02</td>
<td>0.19</td>
<td>0.19</td>
<td>n.s.</td>
<td>-0.01</td>
<td>0.62</td>
<td>0.82</td>
<td>&lt;0.01*</td>
<td>-0.02</td>
<td>0.014*</td>
<td>-0.92</td>
<td>0.04*</td>
<td>-0.04</td>
<td>0.005**</td>
</tr>
<tr>
<td>Proportion FL</td>
<td>0.28</td>
<td>n.s.</td>
<td>-0.02</td>
<td>0.011*</td>
<td>0.20</td>
<td>n.s.</td>
<td>-0.02</td>
<td>0.013*</td>
<td>0.31</td>
<td>n.s.</td>
<td>0.00</td>
<td>0.72</td>
<td>0.93</td>
<td>0.02*</td>
<td>0.00</td>
<td>0.39</td>
</tr>
<tr>
<td>Time Since Injury (Mean)</td>
<td>0.31</td>
<td>0.21</td>
<td>-0.12</td>
<td>0.68</td>
<td>0.11</td>
<td>0.0023**</td>
<td>0.12</td>
<td>0.0726.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time Since Injury (Range)</td>
<td>0.02</td>
<td>0.55</td>
<td>-0.12</td>
<td>0.68</td>
<td>0.01</td>
<td>0.75</td>
<td>0.00</td>
<td>0.89</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Significance codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘n.s.’ not significant

TBI: Traumatic brain injury; FL: Frontal lobe injury; RH: Right hemisphere injury
Subgroup analysis of severity classification of TBI on ToM task performance revealed a larger effect for severe TBIs (ES=1.21, 95% CI [0.87;1.55]) than a mixed severity TBI sample (ES=0.79, 95% CI [0.13;1.44]) across all ToM tasks. For FOTOM and understanding IS, the effect was larger for mixed severity TBIs than only severe TBIs, but these results should be viewed in light of the very small number of studies in the mixed severity subgroup for FOTOM (n=1), SOTOM (n=0) and understanding IS (n=1) tasks.

Table 7

Results from Subgroup Analysis of the Impact of Categorical Variable of Severity of TBI on ToM Task Effect Sizes

<table>
<thead>
<tr>
<th></th>
<th>Cochrane’s Q</th>
<th>Degrees of Freedom</th>
<th>Significance (p)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FOTOM</td>
<td>8.26</td>
<td>2</td>
<td>0.016*</td>
</tr>
<tr>
<td>SOTOM</td>
<td>8.84</td>
<td>1</td>
<td>0.003**</td>
</tr>
<tr>
<td>Understanding IS</td>
<td>11.16</td>
<td>2</td>
<td>0.004**</td>
</tr>
<tr>
<td>Faux Pas</td>
<td>9.53</td>
<td>2</td>
<td>0.009**</td>
</tr>
</tbody>
</table>
Discussion

Summary of Results

The present meta-analysis aimed to investigate the extent of ToM deficits reported in the literature on TBI. The results of the meta-analysis showed medium to large effect sizes, indicating moderate to severe impairment in ToM abilities in individuals who have experienced a TBI. The highest effect size was observed for understanding IS tasks, followed by faux pas, SOTOM and finally FOTOM, however, the effect sizes were associated with overlapping of confidence intervals.

Re-calculation following systematic exclusion of certain studies did not provide evidence of author bias. Overall it was concluded that there is a risk that publication bias may have inflated the observed effect between TBI and performance on FOTOM, SOTOM and understanding IS tasks. However, it is important these findings are interpreted in light of the small number of studies.

The present meta-analysis sought to enhance the homogeneity of the clinical sample by only including studies in which a minimum of 50% of ABIs were traumatic in origin. Sensitivity analysis demonstrated that presence of a higher proportion of the clinical sample with a TBI significantly reduced ToM performance on all tasks except faux pas.

FL location of TBI was found to significantly impact on FOTOM and SOTOM task performance. Location of the injury in the RH had a significant impact on ToM task performance for the relatively more sophisticated tasks of understanding IS and faux pas.

Analysis of the impact of the time since injury indicated that the longer the time between injury and assessment, the poorer the performance of the individual on the more complex tasks of understanding IS and faux pas. The only demographic moderator variable to have a
significant effect on ToM task performance was years in education, which had a significant
effect on faux pas tasks.

*Theoretical Models and Neuroanatomy*

Through inclusion of four ToM tasks, this meta-analysis was able to assess the nature of ToM
impairment using the hierarchy model of ToM development in childhood. Following this
model, it was hypothesized that the skills learned later in childhood would demonstrate the
highest level of impairment following injury. Based on the magnitude of effect sizes, the data
presented in this review do not provide consistent support for this hypothesis. As would be
expected, individuals showed the lowest level of impairment on FOTOM tasks, followed by
SOTOM tasks. This is in-keeping with the hypothesis that less sophisticated skills, which are
learned earlier, remain relatively more preserved. However, comprehension of social faux pas,
which was hypothesized to be the most impaired skill (because it is the last to develop),
showed less impairment than understanding IS. The same pattern of results was reported by
Martín-Rodríguez and León-Carrión (2010), suggesting consistency in this finding. Overall,
the findings indicate low levels of impairment in ToM abilities associated with earlier
development and higher levels of impairment in abilities associated with later development.
Therefore partial support is offered for the proposed hierarchy model of development of ToM
abilities.

Intellectual functioning is associated with ToM ability, with significant correlations between
ToM and vocabulary, information processing speed and executive function reported in the
literature (Maylor, Moulson, Muncer, & Taylor, 2002; Walz, Yeates, Taylor, Stancin, &
Wade, 2009). Due to an absence of data in the published studies, inclusion of intellectual
functioning was not possible in the present analysis. There is a well-established progressive
increase seen in intellectual functioning with an increase in years in education (Crawford & Allan, 1997; Matarazzo & Herman, 1984). As would be expected, the present study found that participants with fewer years of education were more vulnerable to impairment in the complex ToM task of understanding faux pas.

The relationship between intellectual functioning and social cognition is supported by research with patients with temporal lobe epilepsy. Wang et al. (2015) reported that impaired intellectual functioning was a significant predictor of abnormal social functioning. The literature indicates a relationship between estimated IQ scores and some ToM tasks, including FOTOM, SOTOM, cartoon (Hur et al., 2013) and Reading the Mind in the Eyes (Ahmed & Miller, 2011), however, this effect has not previously been observed in faux pas tasks. It may be that ToM tasks share a general attentional capacity which contributes to functioning in these domains (Bull, Phillips, & Conway, 2008; McKinnon & Moscovitch, 2007).

Given that executive functioning increases as educational level increases (Constantinidou, Christodoulou, & Prokopiou, 2012; Dorbath, Hasselhorn, & Titz, 2013), it may have a role in the relationship between intellectual functioning and ToM ability. Some researchers, including Hughes, Russell, and Robbins (1994) argue that ToM tests are themselves tests of executive function. Evidence for this approach comes from autism research, in which individuals with a diagnosis of autism spectrum disorder possess deficits in executive functions (Hughes & Russell, 1993; Hughes et al., 1994; Ozonoff & Mcevoy, 1994).

Unfortunately, in the present study there was insufficient data available to examine the relationship between performance on the ToM tasks and tests of executive functions. Overall, the weight of the body of literature suggests robust correlations between performance on ToM

---

11 A neurodevelopmental disorder in which impaired ToM is believed to explain the core social deficit (McPartland, Reichow, & Volkmar, 2012).
and executive functioning tasks (Carlson, Moses, & Claxton, 2004; Sabbagh, Xu, Carlson, Moses, & Lee, 2006). It is important for models of cognitive processes to clarify the relationship between executive functioning and ToM, and establish the interdependence of these processes. Although meta-analysis may contribute to our understanding of this relationship, this question would be more appropriately addressed by studies using latent variable analysis of multiple ToM tasks.

Impairment in social cognition is recognised as a common outcome of damage to the FL (Price, Daffner, Stowe, & Mesulam, 1990). However, the specific ToM processes and neuroanatomical correlates have remained largely unidentified. The present meta-analysis reported that frontal location of TBI significantly impacted FOTOM and SOTOM task performance. This is consistent with findings from Rowe et al. (2001), who reported impaired ability of individuals with FL lesions to infer first and second order beliefs. The present study cannot, in isolation, illuminate whether the observed deficit in FOTOM and SOTOM abilities occurs solely with frontal lesions as opposed to damage in other brain areas. Nonetheless, the results of the meta-analysis are consistent with localisations identified by functional neuroimaging studies (Fletcher et al., 1995; Goel, Grafman, Sadato, & Hallett, 1995).

Location of the injury in the RH had a significant impact on ToM task performance for the relatively more sophisticated tasks of understanding IS and faux pas. Given that understanding IS and faux pas test pragmatics, it is reasonable to assume that they share brain regions. The reason for the differential effect of FOTOM and SOTOM compared with understanding IS and faux pas may be the role of emotion perception in the latter tasks. Neuropsychological research conducted by Borod (1992) indicated dominance of the RH for emotion perception. The results of the present meta-analysis are consistent with
neuropsychological data, suggesting a crucial role of the RH in more sophisticated ToM abilities which require emotion perception.

Analysis of the impact of time since injury indicated that the longer the time between injury and assessment, the poorer the performance of the individual on the more complex tasks of understanding IS and faux pas. This pattern of findings suggests that comprehension of IS and social faux pas show a deterioration over time. Rather than a purely organic mechanism, the deterioration of complex social processes over time may be attributed to the characteristic social withdrawal and isolation often seen following TBI (Nochi, 1997). This reduction in social activity may mean that individuals have fewer opportunities to strengthen any social skills that may have suffered impairment. It is likely that negative reinforcement would facilitate a cyclical process of avoidance of complex social situations and reduction in ToM ability.

**Methodological Issues**

The literature in this field provides consistent evidence that ToM is a multidimensional construct (Martín-Rodríguez & León-Carrión, 2010; Ubukata et al., 2014), with different tests used to assess different elements. One possible explanation for the observed pattern of results is that the method used to categorise studies into groups based on ToM ability was not accurate. The moderate levels of heterogeneity in SOTOM, faux pas and understanding IS tasks, and extensive overlapping of confidence intervals, may be indicative of dissipation of real differences between TBI and control populations.

The present research sought to extend the previous meta-analysis conducted by Martín-Rodriguez and León-Carrión (2010) which investigated ToM deficits in patients with ABI. The present review attempted to reduce sample heterogeneity and risk of bias through use of a
homogenous TBI sample. Sensitivity analysis revealed that although the sample was relatively homogenous, the presence of a small number of cases of non-traumatic brain injury had an impact on the results. Future studies should endeavour to reduce bias in the clinical sample. Researchers could continue to test this hypothesis through analysing ToM task performance of adult patients who have experienced localised TBI in different stages of childhood. Given the relevance of the FL and RH, it would be of interest to explore the pattern of impairment demonstrated by patients with right-sided FL damage.

Due to an absence of data in the primary studies, the present meta-analysis was unable to investigate the influence of a number of variables related to outcome following brain injury. Few papers reported data on cognitive functioning at time of assessment (memory, attention, executive functioning etc.) or premorbid intellectual functioning. Given the prevailing clinical assumption that higher premorbid intellectual functioning is related to better outcomes following TBI (Kesler, Adams, Blasey, & Bigler, 2003), it follows that it would be important to control for level of intellectual functioning. Unfortunately the sparsity of published data meant this was not possible. Furthermore, neuropsychological studies have indicated that factors other than ToM abilities can contribute to performance on ToM tasks, these include cognitive flexibility (Grafman & Salazar, 2015), executive functions (Channon & Watts, 2003), organisational/planning skills (Grafman & Salazar, 2015) and affective processing (Shamay-Tsoory, Tomer, & Aharon-Peretz, 2005; Shamay-Tsoory et al., 2003; Shamay-Tsoory, Tomer, Berger, et al., 2005). It is a limitation of the present review that these factors were not controlled. Inclusion of these factors in future research would facilitate a more comprehensive understanding of the relationship between TBI and ToM ability.
**Limitations of the Meta-Analysis**

**Small Number of Studies.** Task selection for the present meta-analysis was performed based on a convention set by Martín-Rodríguez and León-Carrión (2010) of inclusion of four commonly used ToM tasks which map onto the developmental model of ToM. Due to the scarcity of other types of ToM task in the literature, studies were excluded if less common types of ToM task were implemented. It is likely that the small number of published studies that utilised lesser-known ToM tasks would have important information to contribute to the field, despite not being included in the present review.

**Verbal and Non-Verbal Material.** An unintended consequence of the task selection criteria was that the tasks included in the meta-analysis tended to use verbal material. It is a limitation of the present study that the similarities and differences in performance across verbal and non-verbal tasks could not be explored further. Neuropsychological evidence of the cognitive consequences of TBI suggests that it is likely that patients would demonstrate poorer performance on verbal tasks than non-verbal tasks, due to the high prevalence of language impairment in TBI (Tennant, Macdermott, & Neary, 1995). However, previous research has not found a significant difference in ToM task performance as a function of verbal or non-verbal material (Martín-Rodríguez & León-Carrión, 2010). Accordingly, it is reasonable to conclude that ToM may not be affected by language deficits and reliance on verbal tasks will have had minimal impact on the overall findings of the present review.

**Clinical Implications**

Impaired ToM capacity can directly interfere with the formation and maintenance of social relationships, as the individual misinterprets the behaviour of others and responds accordingly. There are also distal consequences of TBI, which impact on social functioning.
These include loss of employment and working relationships, social stigma, physical problems\(^{12}\) (Grafman & Salazar, 2015) and of course cognitive difficulties\(^{13}\). These obstacles can lead to social withdrawal and isolation, which may contribute to diminished overall social functioning. By paying attention to the relationship between TBI and ToM ability, the present meta-analysis helps to build a comprehensive picture of the impact of TBI on social and emotional functioning. Improved understanding of these issues is key for the design and implementation of effective rehabilitation programmes.

When carrying out assessments clinical psychologists and clinical neuropsychologists should gather detailed information regarding the injury itself (localisation, lateralisation, age of injury) and type and severity of subsequent ToM deficit. The information gathered in these areas would be an important part of the formulation of an individual’s difficulties, which could be used to inform the development of an individualised treatment package.

The findings of the present meta-analysis suggest that Social Skills Training (SST) groups may be a particularly helpful element of rehabilitation, for individuals recovering from a TBI. The design and implementation of these groups could benefit from more in depth understanding of ToM deficit following TBI and its likely impact on day-to-day, real-world social behaviours (Ylvisaker, Turkstra, & Coelho, 2005).

**Conclusions**

The present meta-analysis provided evidence of positive associations between TBI and impairment in four areas of ToM: first order belief, second order belief, understanding indirect speech and faux pas. There was partial support for a hierarchical model of

---

\(^{12}\) Including light and sound sensitivity, dizziness, fatigue, balance problems and vestibular dysfunction.

\(^{13}\) Including slowed processing speed as well as memory, attention and verbal fluency impairments.
development of TBI abilities in childhood. These findings are in-keeping with the previous meta-analysis conducted by Martín-Rodríguez and León-Carrión (2010), which used a comparatively heterogenous sample. Examination of the social consequences of TBI is a growing area of research and additional published studies are required in order to draw more robust conclusions about the size, nature and underlying mechanisms of the effect of TBI on ToM abilities.
References


https://doi.org/10.1093/scan/nsr071

Leslie, A. M., & Thaiss, L. (1992). Domain specificity in conceptual development: 
https://doi.org/10.1016/0010-0277(92)90013-8

https://doi.org/10.1016/j.neuropsychologia.2010.02.009


Interaction between emotion recognition, mentalizing ability, and social 

https://doi.org/10.1006/brln.1996.0038

McKinnon, M. C., & Moscovitch, M. (2007). Domain-general contributions to social 


Empirical Study

Healthcare Decision Making: How can Cognitive Bias be Reduced to Aid Treatment Choice for Patients with Multiple Sclerosis?
Abstract

Aims
The present study aimed to investigate the factors affecting the treatment decision-making process, and the ability to accept treatment-associated risks, in patients with multiple sclerosis \((n=58)\), non-neurological orthopaedic patients \((n=46)\) and healthy control participants \((n=55)\).

Method
Participants completed a series of three computerised experiments, each consisting of a number of tasks. The experimental tasks reflected real-life clinical situations related to MS and general health conditions as well as social situations. Tasks were systematically varied in content and presentation to manipulate the likelihood of cognitive biases being employed. The role of possible moderator variables was investigated through the measurement and analysis of personality traits, depression, disability level, current treatment (first/second-line therapies) and perception of severity of MS.

Results
Experiment 1 examined whether decision-making is influenced by personal relevance, severity of risks or structure of task content. Findings supported the role of structure of task content, but not severity of risk or personal relevance, on decision-making task performance. Experiment 2 explored whether decision-making is affected by very small risks of highly adverse outcomes (such as death). A significant interaction effect was found between the type of risk, the likelihood of the risk and participant group. Experiment 3 found no support for the previously reported advantage provided by a graphical aid to the understanding of complex probability calculations.

Significant associations were found between neuroticism, level of disability, level of education and personal perception of MS severity on decision-making abilities.

Conclusions
Implications for a Dual Process account of reasoning are discussed and clinical applications of findings for the development of treatment decision aids are identified.
Introduction

Cognitive Biases in Decision-Making

Whilst it is an undisputed fact that humans make errors of reasoning and judgment, it is also true that humans can possess proficiency in rational calculation and logical decision-making. It is this discrepancy in human behaviour that attracts investigation by a wide range of social, medical and scientific disciplines. As put by Birnbaum, Anderson, and Hynan (1990), “many scholars have found it disturbing to think that humans might have been rational enough to invent probability theory but not rational enough to use it in their daily thought” (p. 477).

A dual process account of reasoning has long been recognised, in which a distinction is made between an intuitive mode (Type 1) that relies on rules and heuristics, and an inferential mode of analytical thinking (Type 2) which is more deductive in nature (Croskerry, 2009b; Evans, 2008). In some decision-making circumstances an intuitive approach (Type 1) may be appropriate, whereas in others an analytical approach (Type 2) might be preferred (Simon, 1990). At times a combination of the two may be optimal (Croskerry, 2009b). The respective operating characteristics of the two processes can be seen in Table 1.
Table 1

*Characteristic Properties of the Two Systems in a Dual Process Model of Reasoning:*

*Adapted from (Croskerry, 2005, 2009b; Dawson, 1993; Evans, 2008)*

<table>
<thead>
<tr>
<th>Property</th>
<th>Type 1</th>
<th>Type 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reasoning style</td>
<td><em>Intuitive</em></td>
<td><em>Analytical</em></td>
</tr>
<tr>
<td></td>
<td><em>Heuristic</em></td>
<td><em>Normative</em></td>
</tr>
<tr>
<td></td>
<td><em>Associative</em></td>
<td><em>Deductive</em></td>
</tr>
<tr>
<td></td>
<td><em>Concrete</em></td>
<td><em>Abstract</em></td>
</tr>
<tr>
<td>Speed</td>
<td>Fast</td>
<td>Slow</td>
</tr>
<tr>
<td>Automaticity</td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td>Awareness</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>Propensities</td>
<td>Causal</td>
<td>Statistical</td>
</tr>
<tr>
<td>Vulnerability to bias</td>
<td>High</td>
<td>Lower</td>
</tr>
<tr>
<td>Errors</td>
<td>Common</td>
<td>Few</td>
</tr>
<tr>
<td>Cost</td>
<td>Low</td>
<td>High</td>
</tr>
</tbody>
</table>

The model of diagnostic reasoning developed by Croskerry (2009b) is based on pattern recognition and dual process theory. Although often cited to explain the processes underlying a diagnostic task, Croskerry (2009a) describes his model as ‘universal’ and proposes that it can be applied to decision-making across multiple domains of healthcare. Figure 1 presents the original diagnostic model and its application to decision-making tasks.
Figure 1. Dual process model for diagnostic reasoning applied to decision-making (Croskerry, 2009b). The model is linear, progressing from left to right. The starting point is the presentation of information to the individual. For the purposes of the present study, this would be information based on the risks associated with a medical decision. The pattern of information is either recognised or not. If it is recognised, fast, automatic Type 1 processes are employed. If it is not recognised, Type 2 processes are adopted. Type 2 processes are slower, analytical and less prone to error. Research suggests that both types of processes are influenced by a number of factors, which are shown in italics. A Type 1 response may proceed directly to a diagnosis or decision, or pass through a calibrator before finalisation. Type 2 processes are always subject to calibration before a final decision is made. Both types of processes can be overridden by the other, through rationalisation or dysrationalia. The ‘cognitive miser’ function is predominant, this refers to the tendency to default to the state that requires the fewest cognitive resources (Croskerry, 2009a).
The characteristics of the two types of processes expose them to error in different ways. Most errors occur with Type 1 and may to some extent be expected, whereas Type 2 errors are infrequent and unexpected but may be very consequential when they do occur (Dawson, 1993). Over 40 biases have been described that may impact on reasoning and result in biased decisions (Croskerry, 2009b), these include availability bias\(^\text{15}\) (Tversky & Kahneman, 1973), representative bias\(^\text{16}\) (Kahneman & Tversky, 1972; McCaughey & Bruning, 2010), order effects\(^\text{17}\) (Elstein & Schwarz, 2002), affective error\(^\text{18}\) (Blanchette & Richards, 2010; Kleeberg et al., 2004) and many more. Underlying these biases is the principle that in situations of uncertainty, humans employ a number of simple and quick ‘rules of thumb’. Such cases result in the premature movement from Type 2 analytical processes to the more error prone Type 1 processes. The present study will focus on the biases that are likely to influence healthcare decision-making, within a dual process model. Through understanding these biases, information can be communicated in a way that facilitates the use of Type 2 processes and consequently reduces the risk of error.

**Shared Decision-Making**

The majority of the existing literature looks at how clinicians make treatment decisions, however, 60% of patients want an active or shared role in making decisions about their treatment. Active or shared decision-making when perceived by participants as occurring,

\(^{15}\) The availability bias proposes that if something can be recalled, it must be important, therefore, people tend to heavily weigh their judgments toward more available information (Tversky & Kahneman, 1973).

\(^{16}\) Decision makers form their judgment of an event/target based on the perceived similarity of its characteristics to a pre-existing prototypical category (Kahneman & Tversky, 1972; McCaughey & Bruning, 2010).

\(^{17}\) Information presented later in a case is given more weight (Elstein & Schwarz, 2002).

\(^{18}\) Higher anxiety is associated with risk aversion (lower acceptance of lower probabilities resulting in outcomes of greater value) (Blanchette & Richards, 2010; Kleeberg et al., 2004).
tends to result in improved affective-cognitive outcomes such as understanding, trust, satisfaction, behavioural outcomes such as adherence, treatment decision, health behaviours, and health outcomes such as symptom reduction, quality of life, physiological measures (Shay & Lafata, 2015). Despite the favourable outcomes, research suggests that patients often have poor understanding and recall of quantitative risk information and of the qualitative nature of risks relating to treatment options (Lloyd, 2001). Following this research, the present study will focus on how patients, rather than clinicians can make efficient and unbiased decisions about healthcare.

Recently described as a ‘Prime Time for Shared Decision Making’ (Spatz, Krumholz, & Moulton, 2017), increasing attention is being paid by clinicians and commissioners to the value of shared decision-making. With research suggesting that it can both improve care and reduce costs (Oshima & Emanuel, 2013; Stacey et al., 2014), shared decision-making has been introduced as an alternative to traditional informed consent procedures in some areas (Washington State Legislature, 2007). When adopting a shared decision-making approach, clinicians are expected to use a specialised decision aid to facilitate discussion about different treatment options and their respective risks and benefits. Although this approach has been spearheaded in the United States, the shift towards value-based care in the National Health Service (Gentry & Badrinath, 2017; Porter, 2009; The Economist Intelligence Unit Limited, 2016), alongside growing privatisation of services (Pollock, 2005; Tallis, 2015) means that deeper understanding of shared decision-making is increasingly important in the United Kingdom. The present study has an important part to play in exploring the mechanisms underpinning healthcare decision-making and how these can be used to develop appropriate decision aids.
Cognitive Biases in Healthcare

Studies suggest that more medical errors involve cognitive error than lack of knowledge or information (McGee, 2015). Inappropriate or biased decisions made in these settings are not only frequent (McGee, 2015) but arguably among the most consequential (Croskerry, 2009b). The specific nature of the cognitive biases that emerge in healthcare decisions is largely unknown, but research suggests that individuals show a different pattern of decision-making when the scenario has personal relevance. This is due to a mechanism known as availability bias (Tversky & Kahneman, 1973), which occurs when decision makers form their judgment of the likelihood of an event based on the ease of which similar occurrences can be recalled.

When faced with a decision that involves considering different probabilities of risk information, this bias can result in processing being prematurely moved from Type 2 to Type 1. Consequently, the decision may be influenced by a simplistic cognitive shortcut that fails to consider relevant and potentially critical evidence, and statistical probabilities are erroneously discounted.

Research has demonstrated an availability bias in healthcare settings in a number of ways (Klein, 2005). These include treatment decisions made by professionals who have recently worked with the condition in question (Poses & Anthony, 1991) and patient treatment decisions regarding conditions that receive a lot of media attention (Triplet, 1992; Van Hunsel, Passier, & Van Grootheest, 2009). The influence of an availability bias on performance on gambling tasks has been reported in patients with Multiple Sclerosis (MS). Nagy et al. (2006) implemented an extended version of the Iowa Gambling Test (Bechara, Tranel, & Damasio, 2000), and found that decisions were guided by recent outcome irrespective of gain or loss. In these situations, given the absence of an objective assessment of likelihood, it is reasonable to assume that errorful decisions are made. In order to explore
the possible influence of availability bias in shared decision-making, this study will explore its influence on tasks related to healthcare (including MS), as well as more general real-life scenarios.

There is consensus in the literature that decision-making can be influenced by the way in which a problem is presented or “framed” (Van Schie & Van Der Pligt, 1995). McNeil, Pauker, Sox and Tversky (1982) asked medical practitioners, postgraduate students and outpatients to imagine they had been diagnosed with lung cancer and to choose between two treatments. Different groups of participants received the same information, but its presentation was systematically varied. Across all three populations, the attractiveness of treatments was greater when the treatment information consisted of life expectancy rather than cumulative probability, and when the problem was framed in terms of the probability of living rather than probability of dying. It is proposed that awareness of how statements are framed (positively or negatively) could improve the quality of shared decision-making.

Widely cited research in this area includes Tversky and Kahneman’s (1981) work on Prospect Theory. Prospect theory proposes that when weighing up different options, a loss is more significant than an equivalent gain. In cases of MS, clinical decisions rarely involve equivalent losses and gains like those described in cognitive theory literature. Rather, clinical scenarios are much more likely to involve weighing up the possible risks of treatment options (“losses”) which will either enable them to continue living with the same frequency of relapses, or a reduction in frequency of relapses (“gains”). To explore the impact of framing of risk decisions in MS, the congruence between risk and outcome was manipulated.

The way in which information or data is presented visually also influences how individuals calculate associated probabilities and make decisions (de Bruin, Fischhoff, Millstein, &
Halpern-Felsher, 2000; Gigerenzer & Hoffrage, 1996). Variation in terms of verbal or numerical expressions of probability, use of pictures, spatial reorganisation, Venn diagrams or novel graphical representations can facilitate understanding of the probability relationships and aid problem solving (de Bruin et al., 2000; Gigerenzer & Hoffrage, 1996; Micallef, Dragicevic, & Fekete, 2012; Zahner & Corter, 2010). This simple tool has important implications for assisting patients to make evidence-based healthcare choices, and could be applied to the development of staff training packages, patient information resources and decision-making aids.

**Multiple Sclerosis (MS)**

MS is an inflammatory-demyelinating disease of the central nervous system. Among young adults it is the second most frequent cause of disability (Compston & Coles, 2008). In MS, both the content and process of decision-making is complex. Although disease progression varies, most MS patients are faced with a series of treatment options, in which treatments become more aggressive as the disease progresses. For a number of reasons (continuing disease activity, acute exacerbations, fatigue, depression and medications) there is evidence of cognitive impairment affecting decision-making in MS patients (Burks, Bigley, & Hill, 2009; Chiaravalloti & DeLuca, 2008). MS patients with diminished neuropsychological function show poor understanding of treatment disclosures compared to their neurologically intact peers, and diminished new learning and executive function are correlated with poorer understanding. Nonetheless, with sufficient cuing, MS patients with impaired neuropsychological function are able to display understanding equivalent to their neurologically intact peers (Basso et al., 2008). Accordingly, if shared decision-making is to be achieved the restructuring of information in order to facilitate analytical (Type 2) decision-making may be essential for patients with MS-related cognitive processing difficulties.
Alongside these cognitive impairments, the clinical scenario is made more complicated by the increasing number of new MS drugs that have been developed over the last 10 years and are expected in the near future. These potentially highly effective pharmacotherapies entail risks of potentially serious, although generally infrequent, adverse events (including death). Lichtenstein et al. (1978) reported that individuals show systematic bias in the estimation of frequency of death, both in a tendency to underestimate larger frequencies and overestimate smaller ones, as well as a tendency to exaggerate the frequency of some causes and underestimate the frequency of others. It is reasonable to assume that analytical processing (Type 2) may be overridden when there are small risks of highly adverse outcomes, however, this is yet to be studied in individuals with MS.

Tur et al. (2013) found that acceptance of risk in MS patients was influenced by current treatment (first/second-line therapies)\(^9\) and perception of disease severity, but not level of disability or current functioning. The researchers also investigated the role of the Big Five personality traits (Costa & McCrae, 1992; Goldberg, 1993) which are:

- Extraversion–Introversion
- Neuroticism–Emotional Stability
- Agreeableness–Hostility
- Openness to Experience–Closedness to Experience
- Conscientiousness–Lack of Conscientiousness

\(^9\) First line therapies are treatment regimes that are generally accepted by the medical establishment as the initial treatment for a condition. There are further 'second line' treatments for people who are not responding well or who have a more aggressive form of MS.
Neuroticism was the only personality trait that was associated with risk acceptance. It would be reasonable to expect that individuals with higher levels of neuroticism are more frightened of treatments with adverse effects. However, the authors found that participants presenting with higher levels of neuroticism were actually more prone to assuming higher risks. It could be argued that this was because they were more concerned about the disease progression than the possibility of secondary effects. Notably, more than 50% of MS patients experience depression at some point during the course of their illness (Feinstein & Feinstein, 2001), and in individuals with MS, high neuroticism scores have been associated with higher levels of depression (Bruce & Lynch, 2011). The present study will seek to clarify the roles and possible interactions of depression and personality traits, specifically the neuroticism-emotional stability dimension, on risk decisions.

A limitation of Tur et al.’s (2013) work is that they were unable to draw conclusions regarding whether participants fully understood the contents of the tasks they were presented with. To address this difficulty, Tur et al. (2013) recommended that future research should record and analyse the social and educational status of participants. Accordingly, highest level of education, and current employment/education status will be included as possible moderators of risk calculations and associated decisions.

The Present Study

The present study uses a series of computerised experimental tasks to explore the cognitive biases that may impact on decision-making and ways to overcome them. Healthy controls, non-neurological medical patients and MS patients will be asked to make complex risk decisions. Unlike Tur et al.’s (2013) study, non-neurological medical patients will be included because they present with physical health conditions that should not have significant cognitive consequences. Therefore, they act as a comparison condition to the patients with MS and
healthy controls, aiding the discrimination of the contribution of cognitive impairment to
errors in decision-making.

The content of the intervention will reflect real-life situations, and tasks will be systematically
varied in content and presentation to manipulate the likelihood of cognitive biases. To explore
the likelihood of the availability and personal relevance of information on decision-making,
tasks will involve MS specific, general health or day-to-day/social issues. There will be
systematic variation of the likelihood and severity of adverse effects for several treatment
decisions, including those with a small risk of death. To determine the extent to which
cognitive biases are implemented (and their influence on task outcome), the decisions made
by participants and their risk estimations will be measured as the dependent variables.

Analysis will compare the decisions made within group (patients with MS, non-neurological
medical patients and healthy controls), between groups and across different scenarios (MS
related, general health and neutral topics). It will also investigate whether differential
presentation of information influences the likelihood of cognitive biases being employed and
consequently the decision that is made. There will be exploration of whether personality traits,
mood states, perception of disease severity or current treatment (for those in the MS
condition) influence decision-making. Interpretation of the results will include examination of
the cognitive underpinnings of decision-making in healthy controls and patients with and
without a neurological condition. This has implications for best practice of communicating
treatment options with patients.

The present study aims to investigate the factors affecting the treatment decision-making
process, and the ability to accept treatment-associated risks in patients with MS, non-
neurological patients and healthy control participants.
Hypotheses

Experiment 1: Personal Relevance

1) The understanding and acceptance of risk of adverse events is influenced by personal relevance of the decision.

2) The understanding and acceptance of risk of adverse events is influenced by perception of the severity of adverse outcomes.

3) The understanding and acceptance of risk of adverse events is influenced by the sentence structure of the decision-making task.

Experiment 2: Very Rare Adverse Events

4) The understanding and acceptance of risk of adverse events is influenced by the possibility of very rare and highly adverse events.

Experiment 3: Presentation of Risk Information

5) The understanding and acceptance of risk of adverse events is influenced by whether the information is depicted textually or graphically.

Moderator Effects

6) In addition, differences in the understanding and acceptance of risk of adverse events will be examined according to the following factors, across all participants:

   a) Personality traits
   b) Level of disability
   c) Depression
   d) Highest level of education
   e) Current educational or employment status

And across participants with a diagnosis of MS:

   f) Stage of treatment for MS (current medication)
   g) Level of functioning
   h) Perception of disease severity (of MS in general and personal condition)
Method

Participants

A total of 159 (109 female) participants were recruited, with the control, fracture and MS conditions consisting of 55, 46 and 58 participants respectively. The age of the participants ranged from 18 to 76 years old, with the mean age being 38.7 years. The mean age of participants in each of the control, fracture and MS conditions was 32.3, 42.1 and 42.6 years respectively. Regarding ethnicity, 86% of participants identified as White, with the next most common ethnicities identified as Asian/Asian British (2%) and Mixed/Multiple ethnic groups (2%). 79% participants were in full-time education or employment (See Table 2).

To be included in the study, participants needed to be aged 16 or over and able to read, write and speak English. Participants were excluded from the MS condition if their responsible medical officer indicated that they did not have capacity to make decisions about their treatment, or that their health did not meet the demands of the research. Participants in the control condition and orthopaedic condition were excluded if they had experienced an acquired brain injury (ABI). This was so that the orthopaedic patients could act as a comparison condition to the patients with MS and healthy controls.
# Table 2

**Participant Demographic Information**

<table>
<thead>
<tr>
<th></th>
<th>Control $(n)$</th>
<th>Fracture $(n)$</th>
<th>MS $(n)$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gender</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>44</td>
<td>24</td>
<td>41</td>
</tr>
<tr>
<td>Male</td>
<td>9</td>
<td>22</td>
<td>17</td>
</tr>
<tr>
<td>Data Not Available</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><strong>Ethnicity</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asian/Asian British</td>
<td>3</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>Black/ African/Caribbean/Black British</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Mixed/Multiple Ethnic Groups</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Other Ethnic Group</td>
<td>1</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>White</td>
<td>47</td>
<td>35</td>
<td>55</td>
</tr>
<tr>
<td><strong>Highest Level of Education</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BTEC, NVQ or HNC</td>
<td>3</td>
<td>9</td>
<td>11</td>
</tr>
<tr>
<td>Doctorate/Phd</td>
<td>13</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Masters degree</td>
<td>15</td>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>Other</td>
<td>0</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>Postgraduate diploma or certificate</td>
<td>2</td>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td>Secondary school or equivalent</td>
<td>1</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>Sixth Form/A Levels/O Levels</td>
<td>3</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>Some example answer option</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Undergraduate degree</td>
<td>18</td>
<td>11</td>
<td>13</td>
</tr>
<tr>
<td><strong>Currently in Education or Employment?</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>6</td>
<td>9</td>
<td>19</td>
</tr>
<tr>
<td>Yes</td>
<td>49</td>
<td>37</td>
<td>39</td>
</tr>
</tbody>
</table>
MS Participant Group Information

As described below, participants in the MS group were administered a number of MS specific measures to gather information about their current medication, disease severity and level of functioning (see Appendices G and H for measures in full).

MS Current Medication. As can be seen in Figure 2, most of the participants with MS who took part in the study were receiving first line therapy at the time of participation (38 participants; 65.5%). Equal numbers of participants were either receiving second line therapies and or medication aimed solely at management of symptoms associated with MS (7 participants; 12.1%). The smallest group, of just 6 participants (10.3%), were receiving no medication for MS or associated symptoms at the time of participation. Descriptive statistics for the medication status of participants with MS is presented in Appendix L, Table A1.

![Figure 2. Pie chart presenting MS group medication at time of participation.](image)

Drugs approved in the EU for the treatment of MS as first line therapies include IFN beta 1a (Avonex, Rebif, Plegridy), IFN beta 1b (Betaferon/Betaseron, Extavia), Glatiramer acetate (Copaxone), Dimethylfumarate (Tecfidera) and Teriflunomide (Aubagio). Second line therapies include Natalizumab (Tysabri) and Alemtuzumab (Lemtrada) (Dörr & Paul, 2015). There is some variation in the use of Fingolimod (Gilenya), but following guidance from relevant clinicians at the recruiting medical site, Fingolimod will be classified as a second line therapy for the purposes of this study.

---

20 Drugs approved in the EU for the treatment of MS as first line therapies include IFN beta 1a (Avonex, Rebif, Plegridy), IFN beta 1b (Betaferon/Betaseron, Extavia), Glatiramer acetate (Copaxone), Dimethylfumarate (Tecfidera) and Teriflunomide (Aubagio). Second line therapies include Natalizumab (Tysabri) and Alemtuzumab (Lemtrada) (Dörr & Paul, 2015). There is some variation in the use of Fingolimod (Gilenya), but following guidance from relevant clinicians at the recruiting medical site, Fingolimod will be classified as a second line therapy for the purposes of this study.
**MS Disease Severity.** Patients were asked to rate the extent to which they thought MS was a severe disease (5=less severe; 1=more severe), in general (i.e. for MS patients as a whole) and in their particular case. Therefore, two severity perception scores were obtained per patient: general and personal. Overall, the MS group rated the severity of MS in general as significantly worse than personal current disease condition \(t=6.79, p<.001\) as shown in Figure 3. Descriptive statistics for ratings of personal and general MS severity are presented in Appendix L, Table A2.

![Figure 3](image)

*Figure 3.* Histogram 3a presents MS group ratings of severity of MS in general (for most people); Histogram 3b presents MS group ratings of severity of personal disease condition.

**MS Level of Functioning.** Ratings of individuals’ level of functioning were based on a modified version of the Expanded Disability Status Scale (EDSS; Kurtzke, 1983)\(^{21}\). For participants with MS, the average level of functioning was rated as 3.32 (SD 2.75). This is approximately equivalent to a level of functioning of ‘*some disability but no restrictions in walking*’ (2.0-3.5). The most common level of functioning indicated in the MS group was

\(^{21}\) The Expanded Disability Status Scale (EDSS; Kurtzke, 1983) is described below and presented in full in Appendix H.
‘normal health or no disability’ (0.0), with 18 participants (31.0%) identifying with this level of functioning. At the more severe end of the scale, three participants identified as ‘only walking a few steps, may use wheelchair’ (7.5) and two participants indicated ‘not walking and using wheelchair’ (8.0) (see Appendix L, Table A2). Scores on the EDSS were highly correlated with scores on a measure of general health and disability, which was completed by all participants (detailed below) ($r=0.78$, $p<.01$).

![Histogram of MS group level of functioning based on EDSS score.](image)

Figure 4. Histogram of MS group level of functioning based on EDSS score.

**Recruitment**

Participants with MS were recruited through MS Clinics at Queen Elizabeth Hospital, Birmingham. Participants were identified by their responsible medical officer who determined
whether they met the inclusion criteria for the study and whether their level of health and functioning would meet the demands made by the research.

The non-neurological medical patients were recruited from a Fracture Clinic at the Queen Elizabeth Hospital, Birmingham. Potential participants were given an information sheet and contact details for the chief investigator. The control group were recruited through adverts on research recruitment websites. Students of the University of Birmingham were also recruited through social media, intranet and posters.

For both the MS and orthopaedic control groups, there were no changes to routine practice or procedures. For all participant groups, no external incentives were offered in the recruitment process and it was made clear that participants were under no obligation or pressure to participate.

Procedure

The intervention consisted of a series of computerised questions, hosted on an online survey website. A paper version was also made available according to patient preference. Before the online survey started, participants were presented with an information sheet, explaining the purpose of the study, the nature of involvement and contact details for the research team. Participants were then presented with a consent form, with which agreement was compulsory before continuing with participation. Those participants with a diagnosis of MS were asked a number of questions about their condition. All participants were presented with standardised measures of personality traits, disability and depression as well as demographic questions.

For Experiment 1, participants were randomly allocated to one of two conditions (high or low severity). The questions in the high and low conditions contained the same probabilities but differing severity of risks and adverse outcomes, to allow comparison across groups. In
contrast, all participants completed the same set of decision-making tasks comprising Experiment 2. For Experiment 3, participants were randomly allocated to one of two conditions: textual or graphical presentation of risk information (all participants received one question, with the same probabilities and risk information, which was presented either as text or as a pie chart). Following participation, participants received a debrief sheet and a list of support services.

**Measures**

**MS Specific Measures.** Following the work of Tur et al. (2013), participants in the MS condition were asked a series of questions about their condition (see Appendix G). In order to establish whether they were receiving first or second line therapies (or none at all), patients were asked to list their current medications. A visual analogue scale was implemented for participants to rate their perception of severity of MS as a condition in general (for most people), and their perception of severity of their own illness, from 1 to 5.

**Expanded Disability Status Scale (EDSS; Kurtzke, 1983).** The EDSS is a scale for rating the degree of neurological impairment in MS. The EDSS scale ranges from 0 to 10 in 0.5 unit increments, with each increasing increment representing a higher level of disability. Because scoring is conventionally based on examination by a neurologist, a modified version was used for the purposes of this study. The EDSS was only completed by participants in the MS condition. A recent systematic review of 54 studies reporting the psychometric properties of the EDSS, reported that overall the EDSS had good validity and inter-rater reliability kappa values between 0.32 to 0.76 (Meyer-Moock, Feng, Maeurer, Dippel, & Kohlmann, 2014).

---

22 In the modified EDSS (Kurtzke, 1983) categories were collapsed and descriptions were simplified to enable self-report (see Appendix H). To ensure the measure maintained its accuracy and coherence, the modifications were devised by a Consultant Neurologist who was familiar with the measure.
World Health Organization (WHO) Disability Assessment Schedule 2.0, 12 Item Version (WHODAS 2.0; World Health Organization, 2010). WHODAS is a self-administered assessment instrument developed by the WHO to provide a standardised method for measuring health and disability. The tool captures level of functioning in six domains of life: cognition, mobility, self-care, interacting with others, life activities and community/society participation. Response options range from 1 (no difficulty) to 5 (extreme difficulty or cannot do). WHODAS 2.0 exists in a 36-item version and 12-item version with different options for administration. To reduce the time commitment for participants, the 12-item version was selected for use in the present study. The 36-item version of the WHODAS has been validated in an MS population, and the total scale showed excellent reliability (Cronbach’s $\alpha=0.93$) (Magistrale et al., 2015). Although currently not validated in an MS population, the 12-item version has a Cronbach’s $\alpha$ of 0.94 in participants with Huntington’s Disease (Carlozzi et al., 2015), suggesting good internal consistency in neurological conditions.

Patient Health Questionnaire-9 (PHQ 9; Kroenke, Spitzer, & Williams, 2001). The PHQ-9 is a 9-item module of the PHQ, keyed to DSM-IV major depression criteria. It can be scored categorically to diagnose major and subclinical depressive disorders, but for the purposes of this study, it will be used as a continuous numerical measure to assess the severity of depressive symptoms. The PHQ-9 has been validated across numerous medical and neurological populations (Ferrando et al., 2007). The PHQ-9 has been validated for use in a MS population by a number of studies (Amtmann et al., 2014; Ferrando et al., 2007), including work by Patten et al. (2015) which reported 95% sensitivity and 88.3% specificity.
Ten-Item Personality Inventory (TIPI; Gosling, Rentfrow, & Swann, 2003). The TIPI is a ten-item measure of the Big Five personality dimensions. The TIPI consists of two items for each of the five dimensions. Each item is rated on a 7-point scale that ranges from 1 (disagree strongly) to 7 (agree strongly). Total scores for each subscale range from 2 to 14, with greater scores indicating more pronounced personality traits. The TIPI has been shown to be a valid and reliable measure, with test-retest correlation of $r=0.72$ and Cronbach’s $\alpha$ of 0.68, 0.40, 0.50, 0.73 and 0.45 for the Extraversion, Agreeableness, Conscientiousness, Emotional Stability and Openness to Experience scales respectively. A recent study utilising the TIPI in a MS sample reported that it is a valid measure for use in MS (Strober, Chiaravalloti, Armstrong, & Roberts, 2014).

Decision-Making Questions

Decision-making tasks were developed using standardised classification systems of symptom and condition severity from The Health Utilities Index (Horsman, Furlong, Feeny, & Torrance, 2003), Institute for Medicine (US) (Chvala & Sharfstein, 1999) and NHS Choices (NHS Choices, 2017). A pilot study of the decision-making questions analysed the performance of 24 non-patient participants who were recruited through convenience sampling. The coherence of responses suggested that participants understood the questions and responses were consistent with expectations based on the literature. This suggests that the questions and content were pitched at the right level and as such the decision-making responses were taken as an accurate measure of performance.

---

**Experiment 1: Personal Relevance.** Twenty-four statements were constructed to systematically test whether risk decisions were influenced by personal relevance of the risk, how the information is structured and severity of risks.

Personal relevance of a risk decision was manipulated by including questions about MS treatment risks, treatment risks for a wide range of other medical conditions (e.g. hay fever or heart disease) or a non-medical risk (e.g. buying a new car, accepting a new job). Participants were randomly allocated to one of two conditions, in which they either received high severity (e.g. non-fatal heart attack) or low severity (e.g. itchy skin rash on arms) risks. Building on research into framing effects, the structure in which risk scenarios were phrased was systematically varied. Congruent structure refers to both parts of the statement being framed in the same way, i.e. both positive or both negative, e.g. Treatment B: has a 30% chance of the condition relapsing but also a 20% chance of side effect (this is congruent as both parts are phrased negatively). Incongruent refers to one positive and one negative component, e.g. Treatment B: has a 70% chance of controlling the condition but also a 20% chance of side effect. The operationalisation of the variables manipulated in Experiment 1 is presented in Appendix E and an example question is displayed in Figure 5.
You have ear ache and are faced with the choice between 2 treatments aimed to cure it, would you choose Option A or Option B?

*Choose one of the following answers:*

- Option A has a 20% chance of reducing the severity of your symptoms and a 70% chance of you experiencing chills.
- Option B has a 20% chance of reducing the severity of your symptoms and a 20% chance of you developing a low-grade fever.

*Figure 5.* An example question from Experiment 1, from the low severity condition, with health risk and incongruent format.

**Participant Response and Scoring.** Participants responded by selecting option A or B. Each response was allocated either 1 or 0 points, with 0 points being allocated if the participant selected the response with maximum benefit and minimum risk (mathematically logical response). As such, the higher the overall scores indicate mathematically illogical “incorrect” responses, which minimise benefit and maximise risk; lower scores indicate more “correct” answers.

**Experiment 2: Very Rare Adverse Events and Death.** Twelve questions were developed to systematically assess the understanding and acceptance of risk of adverse events in cases of very unlikely and very adverse possible outcomes, including death. The operationalisation of the variables manipulated in Experiment 2 is presented in Appendix E and an example question is presented below, in Figure 6.
Option A has a 70% chance of improvement in your relapses and a 0.1% chance of death.

Option B has a 20% chance of improvement in your relapses and a 1% chance of death.

Figure 6. An example question from Experiment 2, manipulating the risk type MS and high probability of adverse outcome.

**Participant Response and Scoring.** Responses to Experiment 2 were assessed based on the extent to which the participant indicated they would accept varying risks. Responses were given a score of 0 or 1 with higher scores indicating a preference for minimising risk and lower scores suggesting higher risk acceptance.

**Experiment 3: Presentation of Risk Information.** Participants were randomly allocated to one of two conditions, in which they either received a question presented in text or graphically (see Figures 7 and 8, respectively). The operationalisation of the variables manipulated in Experiment 3 is presented in Appendix E.

At the age of 40, 10% of adults have a medical condition.
Of those, 8% are identified by a routine screening test, but 2% aren’t recognised by the screening test. Of the 90% of adults who don’t have the condition, 9.6% receive an incorrect positive screening test result.
A person in this age group had a positive test result in a routine screening.

What is the probability that they actually have the medical condition?

Please use the sliding scale.

Figure 7. Experiment 3 decision-making question presented as text.
The chart below shows the probability of developing a medical condition for adults at age 40, who participate in routine screening:

Please look at the chart below and answer the following question:

A person in this age group had a positive test result in a routine screening.

What is the probability that they actually have the medical condition?

*Please use the sliding scale.*

0% 100%

*Figure 8. Experiment 3 decision-making question presented graphically.*
**Participant Response and Scoring.** Participants answered by selecting a percentage (0-100%) on a sliding scale. Analysis of these responses was conducted by calculating the difference between the participant’s response and correct answer. Lower scores indicate a smaller difference and thus a more accurate response.

**Analysis Strategy**

**Assumption of Normality.** The skew and kurtosis\(^{24}\) of the dependent variables was calculated and is presented in Appendix M, Table A3. Where non-normal skew or kurtosis was identified, inferential analysis using ANOVA was supplemented with a non-parametric alternative test. Where there is a difference in the conclusion of the parametric and non-parametric test, the most conservative conclusion will be adopted.

**Assumption of Sphericity.** Sphericity\(^{25}\) is an important assumption for analysis of repeated measures. In cases where the assumption of sphericity is violated, the variance calculations may be distorted, which would result in an inflated F-ratio (Field, 2013). In such cases, it is recommended that the ANOVA is appropriately corrected for subsequent analyses, this can be done using the Greenhouse-Geisser (1959) correction.

**Bootstrap Resampling.** Berkovits, Hancock and Nevitt (2000) compared methods for managing the coexistence of non-normality and sphericity. They reported that when normality and sphericity assumptions are simultaneously violated, the most effective way of controlling bias is through use of the bootstrap-\(F\) method. This method relies on random sampling with replacement, assigning measures of accuracy to sample estimates. However, the bootstrap-\(F\)

---

\(^{24}\) If the value of the skew for the data is greater than the two times the standard error of the skew then the distribution is considered to be non-normal. Similarly if the value of the kurtosis is more than twice the standard error this is also an indication that the data are not normal.

\(^{25}\) Sphericity refers to a condition in which the variances of the differences between all possible pairs of within-subject conditions are equal.
method is only appropriate for use with one-way repeated measure ANOVA designs. At present, there is no known modification for the bootstrap method for experimental designs consisting of a factorial ANOVA design with two repeat measures factors and therefore, this preferred method could not be used in the present study. Accordingly, where both non-normality and sphericity co-exist, Greenhouse-Geisser (1959) corrected F tests will be used and parametric tests will be supplemented by the appropriate nonparametric inference test.

Results

Experiment 1: Personal Relevance

A 3 (Risk Type: MS, Health, Social) by 2 (Congruence: Congruent, Incongruent) by 2 (Severity: High, Low) by 3 (Participant Condition: MS, Orthopaedic, Control) mixed ANOVA was conducted, with repeated measures for the risk type and congruence variables.

Sphericity. Application of the Mauchly Test for Sphericity (Mauchly, 1940), indicated that the main effect of risk type violated the sphericity assumption ($W=0.90, \chi^2(2)=16.95, p<.001$). Accordingly, the ANOVA was corrected for subsequent analyses, using the Greenhouse-Geisser (1959) adjustment. The ANOVA table with Greenhouse-Geisser corrected estimates is presented in Table 3 and the marginal means for this design are presented in Figure 9.
Table 3

ANOVA Results with Greenhouse-Geisser Corrected Estimates for Experiment 1

<table>
<thead>
<tr>
<th>Source</th>
<th>Correction</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Risk Type</td>
<td>Sphericity Assumed</td>
<td>22.03</td>
<td>2</td>
<td>11.02</td>
<td>23.00</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>Greenhouse-Geisser</td>
<td>22.03</td>
<td>1.81</td>
<td>12.17</td>
<td>23.00</td>
<td>.000</td>
</tr>
<tr>
<td>Risk Type* Participant Condition</td>
<td>Sphericity Assumed</td>
<td>1.66</td>
<td>4</td>
<td>0.42</td>
<td>0.87</td>
<td>.483</td>
</tr>
<tr>
<td></td>
<td>Greenhouse-Geisser</td>
<td>1.66</td>
<td>3.62</td>
<td>0.46</td>
<td>0.87</td>
<td>.475</td>
</tr>
<tr>
<td>Risk Type* High or Low Severity</td>
<td>Sphericity Assumed</td>
<td>1.48</td>
<td>2</td>
<td>0.74</td>
<td>1.54</td>
<td>.215</td>
</tr>
<tr>
<td></td>
<td>Greenhouse-Geisser</td>
<td>1.48</td>
<td>1.81</td>
<td>0.82</td>
<td>1.54</td>
<td>.217</td>
</tr>
<tr>
<td>Risk Type* Participant Condition* High or Low Severity</td>
<td>Sphericity Assumed</td>
<td>0.69</td>
<td>4</td>
<td>0.17</td>
<td>0.36</td>
<td>.836</td>
</tr>
<tr>
<td></td>
<td>Greenhouse-Geisser</td>
<td>0.69</td>
<td>3.62</td>
<td>0.19</td>
<td>0.36</td>
<td>.817</td>
</tr>
<tr>
<td>Error (Risk Type)4</td>
<td>Sphericity Assumed</td>
<td>148.47</td>
<td>310</td>
<td>0.48</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Greenhouse-Geisser</td>
<td>148.47</td>
<td>280.73</td>
<td>0.53</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Congruence</td>
<td>Sphericity Assumed</td>
<td>8.27</td>
<td>1</td>
<td>8.27</td>
<td>15.28</td>
<td>.000</td>
</tr>
<tr>
<td>Congruence* Participant Condition</td>
<td>Sphericity Assumed</td>
<td>0.96</td>
<td>2</td>
<td>0.48</td>
<td>.89</td>
<td>.412</td>
</tr>
<tr>
<td>Congruence* High or Low Severity</td>
<td>Sphericity Assumed</td>
<td>0.54</td>
<td>1</td>
<td>0.55</td>
<td>1.01</td>
<td>.316</td>
</tr>
<tr>
<td>Congruence* Participant Condition* High or Low Severity</td>
<td>Sphericity Assumed</td>
<td>1.95</td>
<td>2</td>
<td>0.98</td>
<td>1.80</td>
<td>.169</td>
</tr>
<tr>
<td>Error (Congruence)</td>
<td>Sphericity Assumed</td>
<td>83.90</td>
<td>155</td>
<td>0.54</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Risk Type * Congruence</td>
<td>Sphericity Assumed</td>
<td>10.61</td>
<td>2</td>
<td>5.31</td>
<td>19.35</td>
<td>.000</td>
</tr>
<tr>
<td>------------------------</td>
<td>-------------------</td>
<td>-------</td>
<td>---</td>
<td>------</td>
<td>-------</td>
<td>-----</td>
</tr>
<tr>
<td></td>
<td>Greenhouse-Geisser</td>
<td>10.61</td>
<td>1.98</td>
<td>5.36</td>
<td>19.35</td>
<td>.000</td>
</tr>
<tr>
<td>Risk Type * Congruence * Participant Condition</td>
<td>Sphericity Assumed</td>
<td>3.18</td>
<td>4</td>
<td>0.80</td>
<td>2.90</td>
<td>.022</td>
</tr>
<tr>
<td></td>
<td>Greenhouse-Geisser</td>
<td>3.18</td>
<td>3.96</td>
<td>0.80</td>
<td>2.90</td>
<td>.023</td>
</tr>
<tr>
<td>Risk Type * Congruence * High or Low High or Low Allocation</td>
<td>Sphericity Assumed</td>
<td>0.47</td>
<td>2</td>
<td>0.24</td>
<td>0.86</td>
<td>.423</td>
</tr>
<tr>
<td></td>
<td>Greenhouse-Geisser</td>
<td>0.47</td>
<td>1.98</td>
<td>0.24</td>
<td>0.86</td>
<td>.422</td>
</tr>
<tr>
<td>Risk Type * Congruence * Participant Condition * High or Low Allocation</td>
<td>Sphericity Assumed</td>
<td>0.90</td>
<td>4</td>
<td>0.22</td>
<td>0.82</td>
<td>.516</td>
</tr>
<tr>
<td></td>
<td>Greenhouse-Geisser</td>
<td>0.90</td>
<td>3.96</td>
<td>0.23</td>
<td>0.82</td>
<td>.515</td>
</tr>
<tr>
<td>Error (Risk Type*Congruence)</td>
<td>Sphericity Assumed</td>
<td>85.03</td>
<td>310</td>
<td>0.27</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Greenhouse-Geisser</td>
<td>85.03</td>
<td>306.90</td>
<td>0.28</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Participant Condition</td>
<td>13.82</td>
<td>2</td>
<td>6.91</td>
<td>4.91</td>
<td>.009</td>
<td></td>
</tr>
<tr>
<td>High or Low Allocation</td>
<td>2.65</td>
<td>1</td>
<td>2.65</td>
<td>1.88</td>
<td>.172</td>
<td></td>
</tr>
<tr>
<td>Participant Condition * High or Low Allocation</td>
<td>1.61</td>
<td>2</td>
<td>0.81</td>
<td>0.57</td>
<td>.566</td>
<td></td>
</tr>
<tr>
<td>Error</td>
<td>218.39</td>
<td>155</td>
<td>1.41</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Hypothesis 1: The understanding and acceptance of risk of adverse events is influenced by personal relevance of the decision. This hypothesis was assessed by examining the interaction between risk type and participant condition, which was not significant ($F(3.62,306.90)=0.87, p=.475$). This suggests that the personal relevance of a decision topic, based on presence or absence of a current health condition such as MS, did not influence the decisions participants made.

Hypothesis 2: The understanding and acceptance of risk of adverse events is influenced by perception of the severity of adverse outcomes. The influence of severity of outcome was investigated through allocation of questions with either high or low severity of content. The
interaction between severity, condition and risk type was found to be non-significant
\(F(3.62,280.73)=0.36, p=.817\).

**Hypothesis 3: The understanding and acceptance of risk of adverse events is influenced by the sentence structure of the decision-making task.** Whether the question was framed in a congruent or incongruent structure significantly affected decision-making \(F(1,155)=15.28, p<.001\). There was a significant three-way interaction between risk type, congruence and participant condition \(F(3.96,306.90)=2.90, p<.001\) (see Figure 9).

**Non-Parametric Test Comparison.** Because the data is not normally distributed, Friedman’s Test was conducted, which is a non-parametric test, the results of which are presented in Tables A7 and A8 in Appendix P. This non-parametric analysis was statistically significant, confirming the substantive conclusion of the ANOVA \(\chi^2(2)=89.94, p<0.001\).

**Moderator Analyses**

**Personality Traits.** The only personality trait to reveal a significant correlation with responses on the decision-making tasks in this experiment was the Emotional Stability-Neuroticism dimension \(r=0.16, p<.05\). Participants reporting higher levels of neuroticism (and lower levels of emotional stability) gave a poorer performance on the risk calculation tasks, with fewer correct responses.

**Level of Disability.** Overall level of disability, as measured by WHODAS total score, was significantly correlated with responses to congruent questions about MS \(r=0.21, p<.001\) or social issues \(r=0.18, p<.05\). This effect was not observed on incongruent format questions or questions about general health problems.

**Depression.** Self-reported depression, as indicated by PHQ-9 total score, did not significantly affect participant responses to decision-making tasks.
**Educational or Employment Status.** The highest level of education obtained by a participant showed a trend towards a significant effect on the responses given across all decision-making tasks. The data showed that the higher the education level, the more correct answers provided, however, it is important to note that this was not demonstrated to a statistically significant level (Table 4). Whether an individual was in some form of education or employment, or not, at the time of participation did not significantly influence their responses. See Appendix N for full moderator analyses data.

Table 4

*Maximum Likelihood Polyserial Correlation of Highest Level of Education and ToM Tasks*

<table>
<thead>
<tr>
<th>Highest Level of Education</th>
<th>Health Congruent</th>
<th>Health Incongruent</th>
<th>MS Congruent</th>
<th>MS Incongruent</th>
<th>General Congruent</th>
<th>General Incongruent</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.22</td>
<td>0.22</td>
<td>0.32</td>
<td>0.21</td>
<td>0.19</td>
<td>0.17</td>
</tr>
<tr>
<td></td>
<td>.08</td>
<td>.08</td>
<td>.07</td>
<td>.08</td>
<td>0.08</td>
<td>0.08</td>
</tr>
</tbody>
</table>

**Covariate Analysis.** Variables that demonstrated significant correlations with risk decision task performance were carried forward as covariates. Covariate analysis indicated that level of education, disability score and emotional stability affect the means for levels of this analysis. However, these variables do not demonstrate a significant interaction between risk type, congruence and participant condition (see Appendix O, Table A6). Accordingly, the covariates exert a main effect upon the participants’ scores but do not moderate the relationship between risk type, congruence and participant condition.
Moderator Analyses: MS Group

Full correlation and significance information for the MS specific variables in Experiment 1 are presented in Appendix N, Table A5.

Stage of Treatment for MS (Current Medication). Following offerings from previous studies, the influence of first line, second line, symptomatic or no medication on risk decisions was investigated. No significant effect of current medication on risk decisions was found.

Level of Functioning. The effect of current level of functioning (based on EDSS score) revealed a trend towards significance on responses to questions about MS in the congruent structure \( (r=-0.26, p=.053) \), but this was not found to be statistically significant.

Perception of Disease Severity (MS in General and MS Personal Condition). Overall, findings suggested that for individuals with MS, their perception of the severity of the disease in general does not affect their decision-making. In contrast, the perceptions of severity of an individual’s own MS has a significant effect on responses to MS congruent questions \( (r=-0.36, p<.05) \) and the data show a trend towards significance on MS incongruent questions \( (r=0.26, p=.051) \). This effect was not observed for questions about health or social situations.

Experiment 2: Very Rare Adverse Events

A 3 (Risk Type: MS, Health, Social) by 3 (Risk Level: High, Medium, Low) by 3 (Participant Condition: MS, Orthopaedic, Control) mixed ANOVA was performed, with repeated measures on the risk type and risk level variables.

Sphericity. Application of the Mauchly Test for Sphericity (Mauchly, 1940), indicated that the main effects of risk type \( (W=0.82, \chi^2(2)=31.03, p<.001) \) and risk level \( (W=0.94, \chi^2(2)=9.96, p<.05) \) violated the sphericity assumption. Accordingly, the ANOVA was appropriately corrected for subsequent analyses, using the Greenhouse-Geisser (1959) correction.
Table 5

*ANOVA Results with Greenhouse-Geisser Corrected Estimates for Experiment 2*

<table>
<thead>
<tr>
<th>Source</th>
<th>Correction</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Risk Type</td>
<td>Sphericity Assumed</td>
<td>16.12</td>
<td>2</td>
<td>8.06</td>
<td>50.61</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Greenhouse-Geisser</td>
<td>16.12</td>
<td>1.70</td>
<td>9.51</td>
<td>50.61</td>
<td>0</td>
</tr>
<tr>
<td>Risk Type * Participant Condition</td>
<td>Sphericity Assumed</td>
<td>2.99</td>
<td>4</td>
<td>0.75</td>
<td>4.69</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>Greenhouse-Geisser</td>
<td>2.99</td>
<td>3.39</td>
<td>0.88</td>
<td>4.69</td>
<td>0.002</td>
</tr>
<tr>
<td>Error (Risk Type)</td>
<td>Sphericity Assumed</td>
<td>50.33</td>
<td>316</td>
<td>0.16</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Greenhouse-Geisser</td>
<td>50.33</td>
<td>267.94</td>
<td>0.19</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Risk Level</td>
<td>Sphericity Assumed</td>
<td>16.08</td>
<td>2</td>
<td>8.04</td>
<td>34.99</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Greenhouse-Geisser</td>
<td>16.08</td>
<td>1.88</td>
<td>8.53</td>
<td>34.99</td>
<td>0</td>
</tr>
<tr>
<td>Risk Level * Participant Condition</td>
<td>Sphericity Assumed</td>
<td>2.16</td>
<td>4</td>
<td>0.54</td>
<td>2.35</td>
<td>0.054</td>
</tr>
<tr>
<td></td>
<td>Greenhouse-Geisser</td>
<td>2.16</td>
<td>3.77</td>
<td>0.57</td>
<td>2.35</td>
<td>0.058</td>
</tr>
<tr>
<td>Source</td>
<td>Assumptions</td>
<td>df1</td>
<td>df2</td>
<td>Error</td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------------------------------------------</td>
<td>-----------------</td>
<td>-------</td>
<td>-------</td>
<td>------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Error (Risk Level)</td>
<td>Sphericity Assumed</td>
<td>72.61</td>
<td>316</td>
<td>0.23</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Greenhouse-Geisser</td>
<td>72.61</td>
<td>297.70</td>
<td>0.24</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Risk Type * Risk Level</td>
<td>Sphericity Assumed</td>
<td>14.59</td>
<td>4</td>
<td>3.65 22.53 0</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Greenhouse-Geisser</td>
<td>14.59</td>
<td>3.53</td>
<td>4.13 22.53 0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Risk Type * Risk Level * Participant Condition</td>
<td>Sphericity Assumed</td>
<td>4.69</td>
<td>8</td>
<td>0.59 3.62 0</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Greenhouse-Geisser</td>
<td>4.69</td>
<td>7.06</td>
<td>0.66 3.62 0.001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Error (Risk Type*Risk Level)</td>
<td>Sphericity Assumed</td>
<td>102.30</td>
<td>632</td>
<td>0.16</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Greenhouse-Geisser</td>
<td>102.30</td>
<td>557.45</td>
<td>0.18</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Participant Condition</td>
<td></td>
<td>6.21</td>
<td>2</td>
<td>3.10 7.30 .001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Error</td>
<td></td>
<td>67.20</td>
<td>158</td>
<td>0.43</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**Hypothesis 4:** The understanding and acceptance of risk of adverse events is influenced by the possibility of very rare and highly adverse events such as death. This was examined through manipulation of risk level, which demonstrated a significant main effect (F(1.88,297.70)=34.99, p<.001). This effect tells us that regardless of the risk type or participant condition, some risks were still rated significantly differently to others based on their level of likelihood. Analysis also revealed a significant main effect of participant condition (F(2,158)=7.30, p<.05) and risk type (F(1.70,267.94)=50.61, p<.001).

A significant three-way interaction effect was found between risk type, participant condition and risk level (F(7.06,557.45)=3.62, p<.001). This indicates that the interaction among risk type and participant condition varied across the different risk levels, as depicted in Figure 10.

*Figure 10.* Graphical representation of marginal means for Experiment 2.
Non-Parametric Test Comparison. As indicated in Appendix M (Table A3) a number of variables in this analysis were not normally distributed. As such, the data were re-analysed with the non-parametric Friedman’s Test (see Appendix S, Tables A12 & A13), which indicated significance ($\chi^2 (2) = 220.22, p < .001$). This confirms the substantive conclusion of the ANOVA.

Moderator Analyses

Personality Traits. The degree to which participants possessed the personality trait of emotional stability significantly influenced how they rated risks that were low in likelihood and associated with MS ($r = 0.23, p < .001$) or social situations ($r = -0.16, p < .05$). This means that participants reporting higher levels of neuroticism were less willing to accept the risk of death associated with MS treatment than treatment for other severity-matched health conditions. The negative correlation between neuroticism and performance on social decision-making tasks suggest that the higher an individual’s level of neuroticism, the more willing they were to accept the risk of death in a non-health setting. Participant level of conscientiousness was significantly correlated with ratings of risks with low likelihood associated with MS ($r = 0.21, p < .001$). This indicates that the more conscientious a participant was, the more concerned they were with minimising risk.

Level of Disability. The level of disability reported by participants was significantly correlated with responses to questions with low likelihood of risks relating to MS ($r = 0.24, p < .001$) and high likelihood of risks relating to health issues ($r = 0.17, p < .05$).

Depression. No significant effect of level of depressive symptoms was found on risk decisions in this experiment.

Educational or Employment Status. Whether an individual was in education or employment, or not, at time of participation was significantly correlated with their responses to questions
that involved low likelihood of an adverse outcome across MS ($r=-0.19$, $p<.05$) and social domains ($r=-0.20$, $p<.05$). The highest level of education obtained by an individual did not correlate with their responses in this experiment, see Appendix Q.

**Covariate Analysis.** Variables that demonstrated significant correlations with risk decision task performance were carried forward as covariates, but none were found to significantly moderate the effect (see Appendix R, Table A11).

**Moderator Analyses: MS Group**

**Stage of Treatment for MS (Current Medication).** No significant effect of current medication on risk decisions was found in this analysis.

**Level of Functioning.** Current level of functioning (based on EDSS score) showed a significant effect on responses to questions about health problems in the high risk level condition ($r=0.28$, $p<.05$). This suggests that when there is a high risk of an adverse outcome, an individual’s current level of functioning influences their perception of risks related to general health problems.

**Perception of Disease Severity (MS in General and MS Personal Condition).** Participants’ perceptions of the severity of the condition of MS generally or personally did not significantly influence performance. See Appendix Q for full moderator analyses data.

**Experiment 3: Presentation of Risk Information**

A 2 (Data presentation: Graph, Text) by 3 (Participant Condition: MS, Orthopaedic, Control) between-subjects ANOVA was performed. Participants were randomly allocated to one of two conditions, in which they either received the question presented in text or graphically. A total of 149 participants completed Experiment 3; 72 participants received the question presented as a graph and 77 received the question as text (see Table 6).
Hypothesis 5: The understanding and acceptance of risk of adverse events is influenced by whether the information is depicted textually or graphically. As seen in Table 7, question type (graph or text) did not have a significant effect on participants’ responses ($F(1,148)=.23$, $p=.633$). The interaction between participant condition and question type was not significant ($F(2,148)=1.59$, $p=.207$). Therefore, presentation of the information as a graph or text did not significantly affect the risk rating response provided by participants.
Table 7

Tests of Between-Subjects Effects in Experiment 3

<table>
<thead>
<tr>
<th>Source</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Question Presentation</td>
<td>155.89</td>
<td>1</td>
<td>155.89</td>
<td>.23</td>
<td>.633</td>
</tr>
<tr>
<td>Participant Condition</td>
<td>764.51</td>
<td>2</td>
<td>382.26</td>
<td>.56</td>
<td>.573</td>
</tr>
<tr>
<td>Question Presentation *</td>
<td>2171.82</td>
<td>2</td>
<td>1085.91</td>
<td>1.59</td>
<td>.207</td>
</tr>
<tr>
<td>Participant Condition</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Error</td>
<td>97632.65</td>
<td>143</td>
<td>682.75</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>165884.53</td>
<td>149</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Discussion

In recent years, there has been growing interest in the factors affecting treatment decision-making in patients with MS. Research in this area has been conducted alongside the development of new medications, for which treatment regimes require the acceptance of varying degrees of associated risks.

The present study aimed to investigate the factors affecting the treatment decision-making process, and their implications for a dual process account of reasoning, such as that proposed by Croskerry (2009b). Building on previous work by Tur et al. (2013), the performance of MS patients was compared to orthopaedic patients and healthy control participants.
Summary of Findings

Experiment 1. Influenced by the literature base on availability bias, the first experiment in this series investigated whether the understanding and acceptance of risk of adverse events is influenced by personal relevance. Analysis showed that presence or absence of a diagnosis of MS did not influence the decisions participants made regarding MS specific risks. As such, there was no evidence of an influence of personal relevance or availability bias associated with MS in Experiment 1. Severity of risks did not impact on participant performance, but the way in which information was framed did influence decision-making. The data showed that whether a question was framed in a congruent or incongruent structure significantly affected performance on some decision-making tasks.

Moderator analyses in Experiment 1 showed that participants reporting lower levels of emotional stability (higher neuroticism) provided fewer correct risk calculations. In addition, participant level of disability influenced responses to congruent questions about MS or social issues. The higher the education level of participants, the more correct answers were provided, across all decision-making tasks in Experiment 1. For the group of participants who had a diagnosis of MS, perception of their own health condition as severe was associated with higher acceptance of treatment-related risks.

Experiment 2. The second experiment explored the presence of cognitive biases when participants are asked to make decisions regarding small risks of very adverse events, including death. Participants rated risks differently based on their level of likelihood, and a significant interaction effect was found between the type of risk, the likelihood of the risk and participant group. Therefore, the data supported the hypothesis that the understanding and
acceptance of risk of death is influenced by the possibility of very rare and highly adverse events.

Mirroring the findings of Experiment 1, the degree to which participants reported emotional stability influenced how they rated risks. This was only observed for risks that were low in likelihood and associated with MS or social situations. Furthermore, participant level of conscientiousness was also associated with ratings of low likelihood risks regarding MS. The level of disability reported by participants was associated with calculation of low likelihood of risks relating to MS and high likelihood of risks relating to health issues. Whether an individual was in education or employment, or not, at the time of participation was significantly correlated with their responses to questions that involved low likelihood of an adverse outcome across MS social domains. For patients with MS, the data suggest that when there is a high risk of an adverse outcome, individuals with poorer functioning are more likely to accept greater risks in the pursuit of an effective treatment. However, this was not demonstrated to a statistically significant level.

**Experiment 3.** It was hypothesised that participants’ understanding and acceptance of risk of adverse events would be influenced by whether the information is depicted textually or graphically. Experiment 3 tested this hypothesis by presenting participants with the same risk calculation scenario, presented in either text or a pie chart. The analysis revealed that presentation of information as a text or graph did influence how participants rated risks.

**Theoretical Implications**

This study investigated the role of cognitive biases in healthcare decision-making and provided partial support for a dual process account of reasoning, such as the universal model proposed by Croskerry (2009a, 2009b).
Existing models of decision-making tend to focus on the influence of cognitive or situational variables on human behaviour, with emotional components rarely addressed (Gutnik, Hakimzada, Yoskowitz, & Patel, 2006). Given the interconnections within the brain and the commonality of pathways, it is reasonable to assume the presence of some interplay of emotion and cognition in decision-making. Croskerry’s (2009a, 2009b) dual process model explains that both Type 1 and Type 2 processes have pertinent emotional elements (highlighted in Figure 11). Type 1 processes are more influenced by affective state, whereas Type 2 processes incorporate emotion-related associations, intuitions and past experiences in decision-making. Further evidence for a dual process approach is provided by Kahneman (2012), who from a recent review of various experiments, concluded that there is “growing evidence that good mood, intuition, creativity, gullibility, and increased reliance on System (Type) 1 form a cluster. At the other pole, sadness, vigilance, suspicion, an analytic approach, and increased effort also go together” (p. 69).
Figure 11. Dual process model for diagnostic reasoning and decision-making with emotional elements highlighted (Croskerry, 2009b).
The decision-making tasks in the present study appeared to require predominantly Type 2 processes (hypothetico-deductive reasoning and probabilistic reasoning). It was found that participants with higher levels of neuroticism (and lower levels of emotional stability) struggled with these tasks and were less willing to accept adverse risks, even if they were highly unlikely. This finding does not seem to support the dual process model or Kahneman’s (2012) assertion that sadness (or emotional distress) is associated with less errorful Type 2 processing.

A possible explanation for this seeming contradiction between theory and results is the distinction between state (temporary and dependent on current situation) and trait (long-standing and stable characteristic) levels of personality variables. The personality measure used in the present study was concerned with long-standing trait levels of neuroticism over time. In contrast, the dual process model relies on emotional and personality variables at the specific time that the task was completed. Future studies would benefit from including manipulation and measurement of state emotional variables, to explore the impact of current affective state on decision-making.

The present study expected to find evidence of a personal relevance bias across all participants in the MS condition, when faced with complex risk decisions regarding MS. However, an influential effect of personal relevance on risk calculation and acceptance was only observed in those who perceived their own MS condition as severe. It is argued that this provides evidence for an availability bias, in which individuals with more severe perceptions of their MS have easier cognitive access to occurrences of highly unpleasant treatment effects and/or symptoms. Based on the assumption that occurrences which are more available are more important (Tversky & Kahneman, 1973), these individuals are more likely to weigh risks disproportionately, consequently making biased decisions. This can be understood in a
dual process framework, in which Type 2 analytical processes are discounted in favour of
Type 1 intuitive processes.

Clinical Implications

The findings of Experiment 1 and 2 indicate that the framing of risk information and presence
of small risks of death can influence an individual’s calculation of, and willingness to accept,
risk. This finding is consistent with that observed by Lichtenstein et al. (1978), who reported
that individuals show systematic bias in the estimation of frequency of death. Applying a dual
process model, it is reasonable to assume that analytical processing (Type 2) may be
overridden when there are small risks of highly adverse outcomes. A greater understanding of
the healthcare decision-making process has important implications for patients, families,
clinicians, policy-makers and commissioners.

This finding has important practical applications for the development of decision aids used to
support shared decision-making. Specialised decision aids aim to facilitate discussion about
different treatment options and their respective risks and benefits. Systematic review of the
literature reveals that few decision aids are available for use in MS and they vary in quality
(Edwards & Elwyn, 2009). A Cochrane Review (Stacey et al., 2014) of studies including
more than 30 000 patients showed that when patients use decision aids as part of shared
decision-making, they are more knowledgeable about the options; have more accurate
expectations about the risks and benefits; and are more likely to make decisions that are
consistent with their values, preferences and goals. From their recent review of the literature,
Spatz et al. (2017) concluded that it is imperative that clinicians, patients and policy-makers
are confident that the decision aids used in routine clinical practice are evidence-based,
balanced and meet patients’ needs. Due to the variability in significance of results in the
present study, conclusions drawn are tentative. It is recommended that researchers and
clinicians involved in the design and implementation of decision aids carefully consider how information is structured, as well as the severity with which a patient perceives their condition. This would facilitate a healthcare decision-making process that is not only safer for staff, but provides a more empowering patient experience.

The literature suggests that decision-making in healthcare is changing, as the NHS moves towards a shared decision-making approach, in a context of increasingly value-based care (Spatz, Krumholz, & Moulton, 2017; Gentry & Badrinath, 2017; Porter, 2009; The Economist Intelligence Unit Limited, 2016). As such, it is crucial that professionals involved in policy, commissioning and strategy for MS services have an understanding of the relevant factors and processes involved when patients with MS make treatment decisions. This would facilitate the operation of a health service that not only meets the needs of the people it serves, but is better value for money.

**Limitations and Future Research**

Given the paucity of standardised tools for measuring acceptance of risk in healthcare decision-making, the questions used in the present study were designed specifically to address the hypotheses being tested. As such, the decision-making tasks used had not been independently validated. Despite being embedded in the existing literature base, and tested with a pilot study, it is possible that the decision-making questions were not an accurate measure of performance. In addition, a slightly modified version of the EDSS was used in this study, which although not formally validated, was developed under the guidance of a Consultant Neurologist. Nonetheless, further research is required to assess the validity of the measures used in the present study across populations.
The higher the education level of participants, the more correct answers were provided, across all decision-making tasks in Experiment 1, but this factor had no influence on performance in Experiment 2. This may mean that in Experiment 1, some participants did not fully understand the content of the tasks they were faced with, limiting the extent to which results can be generalised. Alternatively, it may be that individuals with a higher level of education were less vulnerable to the influence of cognitive biases, perhaps due to relatively more reliance on Type 2 processes, which are less prone to error. This fits with the universal model of diagnostic reasoning developed by Croskerry (2009a, 2009b) (see Figure 1), which proposes that education and intellectual functioning facilitate Type 2 processing.

Although information was gathered on an individual’s current and historical educational and employment status, there is a risk that current cognitive impairment remained uncontrolled for by the present study. Participants did not undergo cognitive assessment and the present study did not include brain imaging. Therefore, particularly for the MS participant group, whether the brain pathways involved in making risk decisions were sufficiently undamaged cannot be known. Accordingly, it is proposed that future studies of decision-making in healthcare should include at least a brief cognitive screen, and if possible, comprehensive neuropsychological assessment including imaging, in order to reduce the risk of bias from possible brain damage and/or cognitive impairment.

Despite some evidence for an influential role, evidence of cognitive biases was not observed in the hypothesised manner in the present study. This may be because the study content was not sufficiently powerful to initiate the biases to a magnitude in which Type 2 analytical processes would be impacted. Specifically, hypothetical scenarios in which the participant was asked to imagine the presence of a medical condition may not be similar enough to real-
life clinical situations. It is proposed that the use of more immersive methodologies (such as detailed vignettes, role-play or guided imagery) is more likely to trigger cognitive biases.

**Conclusions**

This study investigated a number of cognitive biases, and found evidence for a personal relevance bias for individuals who perceive their MS severely, when faced with hypothetical MS decision tasks. The study suggests that the phrasing and sentence structure of risk information influences how it is interpreted.

Several personality and pathology variables were found to be associated with decision-making in risk scenarios. Neuroticism, level of education, current education or employment factors and level of disability were related to various types of decision-making across participant groups.

The results of the present study are consistent with a dual process model of reasoning, such as that proposed by Croskerry (2009a, 2009b). There are important clinical implications for patients and professionals, with the potential to enhance confidence and satisfaction with decision-making as well as improve clinical and financial outcomes. The findings of the present study can make an important contribution to the development of high quality MS-specific decision aids to support this patient group in their healthcare journey.
References


https://doi.org/10.1016/S0092-6566(03)00046-1


120


Executive Summary: Public Domain Briefing Document

The Effect of Acquired Brain Injury on Theory of Mind and Decision Making
THE EFFECT OF ACQUIRED BRAIN INJURY ON THEORY OF MIND AND DECISION MAKING

Executive Summary: Public Domain Briefing Document

This document provides an overview of the thesis submitted in partial fulfilment of the requirements of the degree of Doctor of Clinical Psychology (D.Clin.Psy) from the University of Birmingham. The thesis has two parts, a literature review and a research study, which are summarised below.

Literature Review: Theory of Mind Impairment in Individuals who have Experienced Traumatic Brain Injury: A Meta-Analysis

Background. Theory of mind (ToM) reasoning is the ability to infer or reflect on the content of one’s own and other’s mental states (Premack & Woodruff, 1978) and is essential for effective social functioning (Aboulafia-Brakha, Christe, Martory, & Annoni, 2011). Studies have shown that individuals who have experienced a Traumatic Brain Injury (TBI) show impairments in various components of social cognition, including ToM (McDonald & Flanagan, 2004; Milders, Fuchs, & Crawford, 2003). These changes post-injury are strongly associated with problems in daily life and unfavourable longer-term outcomes (Crépeau & Scherzer, 1993; Felmingham, Baguley, & Crooks, 2001; Vilkki et al., 1994). Better understanding of the nature and severity of ToM deficit following TBI has important implications for rehabilitation planning and care management.
There are a number of different tests for measuring ToM abilities, but the four most commonly used are: first-order belief (FOTOM) task, second order belief (SOTOM) task, understanding indirect speech (IS) and social faux pas. These tasks map onto a hierarchical model of development of ToM abilities in childhood, shown in Figure 1. In the present study, it was hypothesized that participants are most likely to demonstrate largest impairment in faux pas, then understanding IS, then SOTOM, and then finally, FOTOM tasks.

![Figure 1](image.png)

*Figure 1. The development of ToM abilities in childhood.*

**Aim.** This review aimed to extend the current literature, by investigating the extent of ToM deficits reported in the literature on TBI. It also examined factors that could explain the differences in ToM performance, such as sample characteristics, methodological features and clinical information related to the TBI.

**Method.** Research databases were systematically searched in order to identify all of the relevant studies of ToM in individuals who have experienced a TBI. In total the search identified 19 studies, which included 28 comparisons of TBI and control group performance on ToM tasks. A statistical procedure known as meta-analysis was carried out, which combined the data from the 19 studies.
To identify and assess any risks of bias in the studies, a quality framework was developed and specifically tailored to the individual needs of this study. The quality framework assessed risk of bias in five domains: Selection Bias, Methodological Bias, Measurement Bias, Statistical Bias and Reporting Bias.

**Results.** The results of the meta-analysis showed medium to large effect sizes, this means that individuals who have experienced a TBI show moderate to severe impairment in ToM abilities. The largest effect was observed in understanding IS tasks, followed by faux pas, SOTOM and finally FOTOM.

The severity of impairment shown on ToM tasks was influenced by:

- The location of the injury (particularly frontal lobe or right-sided),
- Whether the injury had a traumatic origin,
- The length of time between the injury and testing and
- Level of education.

**Conclusions.** The pattern of results provides partial support for a hierarchy model of ToM development in childhood. Findings demonstrate consistency with frontal lobe localisation of basic ToM abilities (FOTOM and SOTOM) and right-sided lateralisation of more complex ToM abilities (faux pas and understanding indirect speech). This research has important implications for the design and delivery of neuropsychological rehabilitation packages.
Research Study: Healthcare Decision Making: How can Cognitive Bias be Reduced to Aid Treatment Choice for Patients with Multiple Sclerosis?

Background. Studies suggest that more medical errors involve cognitive error than lack of knowledge or information (McGee, 2015). Inappropriate or biased decisions made in these settings are not only frequent but arguably among the most consequential (Croskerry, 2009; McGee, 2015). The majority of the existing literature looks at how clinicians make treatment decisions, however, 60% of patients want an active or shared role in making decisions about their treatment. Active or shared decision-making when perceived by participants as occurring, tends to result in better psychological outcomes (such as trust and satisfaction), improved health behaviours and higher quality of life (Shay & Lafata, 2015).

Multiple Sclerosis (MS) is a disease of the central nervous system and is the second most frequent cause of disability in young adults (Compston & Coles, 2008). In MS, both the process and content of decision-making is complex. Although disease progression varies, most MS patients are faced with a series of treatment options, in which treatments become more aggressive as the disease progresses. For a number of reasons (continuing disease activity, acute exacerbations, fatigue, depression and medications) there is evidence of cognitive impairment affecting decision-making in MS patients (Burks, Bigley, & Hill, 2009; Chiaravalloti & DeLuca, 2008).

The processes involved in making treatment decisions can be understood in a dual process framework, such as that proposed by Croskerry (2009). In Croskerry’s model, there are two types of reasoning (Type 1 and Type 2). The different characteristics of the two types of processes expose them to error in different ways (see Table 1). Most errors occur with Type 1 and may to some extent be expected, whereas Type 2 errors are infrequent and unexpected but may be very consequential when they do occur (Dawson, 1993). Over 40 biases have been
described that may impact on reasoning and result in biased decisions (Croskerry, 2009). The present study focuses on the influence of the following potential biases:

- Availability bias (Tversky & Kahneman, 1973) and personal relevance of risk information,
- Perception of the severity of adverse outcomes,
- Framing effects sentence structure of the decision-making task (McNeil, Pauker, Sox & Tversky, 1982; Tversky & Kahneman, 1981; Van Schie & Van Der Pligt, 1995),
- The possibility of very rare and highly adverse events, such as death, and
- Whether information is depicted textually or graphically.

In addition, differences in the understanding and acceptance of risk of adverse events was examined according to the following factors, across all participants:

- Personality traits
- Level of disability
- Depression
- Educational or employment status

And across participants with a diagnosis of MS:

- Stage of treatment for MS (current medication)
- Level of functioning
- Perception of disease severity (MS in general and MS personal condition)
**Table 1**

*Characteristic Properties of the Two Systems in a Dual Process Model of Reasoning:*

*Adapted from (Croskerry, 2005, 2009; Dawson, 1993; Evans, 2008)*

<table>
<thead>
<tr>
<th>Property</th>
<th>Type 1</th>
<th>Type 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reasoning style</td>
<td><em>Intuitive</em></td>
<td><em>Analytical</em></td>
</tr>
<tr>
<td></td>
<td>Heuristic</td>
<td>Normative</td>
</tr>
<tr>
<td></td>
<td>Associative</td>
<td>Deductive</td>
</tr>
<tr>
<td></td>
<td>Concrete</td>
<td>Abstract</td>
</tr>
<tr>
<td>Speed</td>
<td>Fast</td>
<td>Slow</td>
</tr>
<tr>
<td>Automaticity</td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td>Awareness</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>Action</td>
<td>Reflexive, skill-based</td>
<td>Deliberate, rule-based</td>
</tr>
<tr>
<td>Propensities</td>
<td>Causal</td>
<td>Statistical</td>
</tr>
<tr>
<td>Vulnerability to bias</td>
<td>High</td>
<td>Lower</td>
</tr>
<tr>
<td>Errors</td>
<td>Common</td>
<td>Few</td>
</tr>
<tr>
<td>Cost</td>
<td>Low</td>
<td>High</td>
</tr>
</tbody>
</table>

*Aim.* The present study aimed to investigate factors affecting the treatment decision-making process, and the likelihood of biased decisions, in 58 multiple sclerosis patients, 46 non-neurological orthopaedic patients and 55 healthy control participants.

*Method.* Participants completed a series of 3 computerised experiments, each consisting of a number of tasks. The experimental tasks reflected real-life clinical situations related to MS and general health conditions as well as social situations. Tasks were systematically varied in
content and presentation to manipulate the likelihood of cognitive biases. Participants also completed measures of the following factors: personality traits, depression, disability level, current treatment and perception of severity of MS.

**Results.** Findings from Experiment 1 supported the role of structure of task content, but not severity of risk or personal relevance, on decision-making task performance. Experiment 2 tested the hypothesis that decision-making is affected by very small risks of highly adverse outcomes such as death. A significant interaction effect was found between the type of risk, the likelihood of the risk and participant group. Experiment 3 found no support for the hypothesis that whether risk information is presented in text or graphics influences the calculation made.

Significant associations were found between neuroticism, level of disability, level of education and perception of an individual’s severity of MS on decision-making abilities.

**Conclusions.** The findings provide partial support for a dual process account of reasoning. Findings have important practical applications for the development of specialised decision aids used to support shared decision-making.
References


Appendix A

Participant Recruitment Poster

(Document Redacted)
Appendix B

Information Sheet

(Document Redacted)
Appendix C

Consent Sheet

*I confirm that I have understood the information for the study. I have had the opportunity to consider the information, contact the team to ask questions and to have these questions answered satisfactorily.

Please type your initials (of your first and second name) into the box to indicate you have understood and agree.

*I understand that my participation is voluntary and that I can stop at any time, without giving any reason, without my own or my loved one’s medical/social care or legal rights being affected.

Please type your initials (of your first and second name) into the box to indicate you have understood and agree.

*I understand that once I have submitted my responses, it will not be possible to withdraw from the study, unless I contact the research team with my ID code for this study and ask for my data to be withdrawn.

Please type your initials (of your first and second name) into the box to indicate you have understood and agree.

*I understand that the data collected during this study will be looked at by the researcher and relevant others at the University of Birmingham to ensure that the analysis is a fair and reasonable representation of the data.

Please type your initials (of your first and second name) into the box to indicate you have understood and agree.

*I agree to take part in the above study.

Please type your initials (of your first and second name) into the box to indicate you have understood and agree.
Appendix D

Debrief Sheet with Support Services and Contact Details

(Document Redacted)
Appendix E

Operationalisation of Variables

**Experiment 1**

<table>
<thead>
<tr>
<th>Risk Type</th>
<th>Topic or condition of the question</th>
<th>MS, Health, Social</th>
</tr>
</thead>
<tbody>
<tr>
<td>Within Subjects</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Congruence</td>
<td>Structure of the question</td>
<td>Congruent, Incongruent</td>
</tr>
<tr>
<td>Within Subjects</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Severity</td>
<td>Level of severity of adverse</td>
<td>High, Low</td>
</tr>
<tr>
<td>Between Subjects</td>
<td>outcomes of decision</td>
<td></td>
</tr>
<tr>
<td>Personal Relevance</td>
<td>Participant condition</td>
<td>MS, Orthopaedic, Control</td>
</tr>
<tr>
<td>Between Subjects</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Experiment 2**

<table>
<thead>
<tr>
<th>Risk Type</th>
<th>Topic or condition of the question</th>
<th>MS, Health, Social</th>
</tr>
</thead>
<tbody>
<tr>
<td>Within Subjects</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Risk Level</td>
<td>Different probabilities of likelihood of outcome</td>
<td>High, Medium, Low</td>
</tr>
<tr>
<td>Within Subjects</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Participant Condition</td>
<td>Neurological condition, medical condition or healthy</td>
<td>MS, Orthopaedic, Control</td>
</tr>
<tr>
<td>Between Subjects</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Experiment 3**

<table>
<thead>
<tr>
<th>Participant Condition</th>
<th>Neurological condition, medical condition or healthy</th>
<th>MS, Orthopaedic, Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between Subjects</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data Presentation</td>
<td>The same risk question was presented as either a pie chart or text with numerical probabilities</td>
<td>Graph, Text</td>
</tr>
<tr>
<td>Between Subjects</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Examples of Decision-Making Questions

**Experiment 1**

You have ear ache and are faced with the choice between 2 treatments aimed to cure it, would you choose Option A or Option B?

*Choose one of the following answers:*

- Option A has a 20% chance of reducing the severity of your symptoms and a 70% chance of you experiencing chills.
- Option B has a 20% chance of reducing the severity of your symptoms and a 20% chance of you developing a low-grade fever.

*Figure A1.* An example question from Experiment 1, from the low severity condition, with health risk and incongruent format.

You have heart disease and are faced with the choice between 2 treatments, would you choose Option A or Option B?

*Choose one of the following answers:*

- Option A has a 70% chance of the same frequency of your symptoms and a 70% chance of developing Type II Diabetes.
- Option B has a 70% chance of the same frequency of your symptoms and a 20% chance of developing a non fatal heart attack.

*Figure A2.* An example question from Experiment 1, from the high severity condition, with health risk and congruent format.
You have MS and are faced with the choice between 2 treatments, do you choose Option A or Option B?
Choose one of the following answers

- Option A has a 20% chance of reducing the frequency of relapses and a 70% chance of severe pain that prevents most activities.
- Option B has a 20% chance of reducing the frequency of relapses and a 20% chance of severe trouble breathing.

“Multiple Sclerosis (also known as MS) is a disease of the central nervous system where problems occur as episodes of dysfunction such as weakness, tingling, pins and needles, blurring of vision or double vision which may last from a few days to several weeks. These will often improve or even resolve completely. These episodes occur with varied frequency and are unpredictable. These episodes are called relapses and occur in relapsing remitting multiple sclerosis. Sometimes disability in MS can change gradually and this is known as progressive MS. There are treatments for MS that can reduce the frequency of relapse episodes, but do not alter progressive MS.”

*Figure A3.* An example question from Experiment 1, from the high severity condition, with MS risk and incongruent format.

You have MS and are faced with the choice between 2 treatments, would you choose Option A or Option B?
Choose one of the following answers

- Option A has a 70% chance of the same frequency of relapses and a 20% chance of developing an itchy skin rash on arms.
- Option B has a 70% chance of the same frequency of relapses and a 70% chance of developing an itchy skin rash on legs.

*Figure A4.* An example question from Experiment 1, from the low severity condition, with MS risk and congruent format.
You see a colleague in your office drop their litter on the floor, do you choose Option A or Option B?
Choose one of the following answers

- Option A: confronting them has a 70% chance of reducing the frequency of their behaviour and a 70% chance of them punching you.
- Option B: ignoring them has a 70% chance of reducing the frequency of their behaviour and a 20% chance of them punching you.

*Figure A5.* An example question from Experiment 1, from the high severity condition, with general risk and incongruent format.

You are upgrading your car and expect the new car to perform better. You have found 2 cars which you are considering buying. Based on the information provided below, which car would you choose?

Choose one of the following answers

- Car A has a 20% chance of having the same performance as your previous car and a 70% chance of breaking down and needing a lot of repair work.
- Car B has a 20% chance of having the same performance as your previous car and a 20% chance of breaking down and needing a lot of repair work.

*Figure A6.* An example question from Experiment 1, from the low severity condition, with general risk and congruent format.
**Experiment 2**

You have MS (multiple sclerosis) and are faced with the choice between 2 treatments to reduce (improve) relapses, do you choose Option A or Option B?

*Choose one of the following answers:*

- Option A has a 70% chance of improvement in your relapses and a 5% chance of death.
- Option B has a 20% chance of improvement in your relapses and a 0.5% chance of death.

*Figure A7. An example question from Experiment 2, manipulating the risk type MS and high probability of adverse outcome.*

You have type 2 diabetes and are faced with the choice between 2 treatments, do you choose Option A or Option B?

*Choose one of the following answers:*

- Option A has a 70% chance of improvement in your symptoms and a 0.01% chance of death.
- Option B has a 20% chance of improvement in your symptoms and a 0.001% chance of death.

*Figure A8. An example question from Experiment 2, manipulating the risk type health and low probability of adverse outcome.*

You are upgrading your car and expect the new car to perform better. You have found 2 cars which you are considering buying. Based on the information provided below, which car would you choose?

*Choose one of the following answers:*

- Option A has a 70% chance of improvement in your relapses and a 5% chance of death.
- Option B has a 20% chance of improvement in your relapses and a 0.5% chance of death.
Option A has a 20% chance of better performance than your previous car and a 0.1% chance of death due to a car accident.

Option B has a 70% chance of better performance than your previous car and a 1% chance of death due to a car accident.

Figure A9. An example question from Experiment 2, manipulating general risks and medium probability of adverse outcome.

Experiment 3

At the age of 40, 10% of adults have a medical condition. Of those, 8% are identified by a routine screening test, but 2% aren’t recognised by the screening test. Of the 90% of adults who don’t have the condition, 9.6% receive an incorrect positive screening test result.

A person in this age group had a positive test result in a routine screening.

What is the probability that they actually have the medical condition?

Please use the sliding scale.

Figure A10. Experiment 3 decision-making question in presented in text.
The chart below shows the probability of developing a medical condition for adults at age 40, who participate in routine screening:

Please look at the chart below and answer the following question:

A person in this age group had a positive test result in a routine screening. What is the probability that they actually have the medical condition?

*Figure A11. Experiment 3 decision-making question in presented in text.*
Appendix G

MS Specific Measures

Please could you let us know any medications you are currently taking for your multiple sclerosis (MS)?

How severe do you perceive your multiple sclerosis (MS) to be at present?

1 = more severe
6 = less severe

How severe do you perceive the condition of multiple sclerosis (MS) to be in general for most people?

1 = more severe
6 = less severe
Appendix H

Expanded Disability Status Scale (EDSS; Kurtzke, 1983) (Survey Version)

(Document Redacted)
Appendix I

World Health Organization (WHO) Disability Assessment Schedule 2.0, 12 Item Version (WHODAS 2.0; World Health Organization, 2010)

(Document Redacted)
(Document Redacted)
Appendix J

Patient Health Questionnaire-9 (PHQ 9; Kroenke et al., 2001)

(Document Redacted)
Appendix K

Ten-Item Personality Inventory (TIPI; Gosling, Rentfrow, & Swann, 2003)

(Document Redacted)
Appendix L

MS Group Variables Data

Table A1

*MS Group Current Medication Information*

<table>
<thead>
<tr>
<th>Number of Participants</th>
<th>Percentage of Participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>(n=58)</td>
<td></td>
</tr>
<tr>
<td>No Medication</td>
<td>6</td>
</tr>
<tr>
<td>First Line Therapy</td>
<td>38</td>
</tr>
<tr>
<td>Second Line Therapy</td>
<td>7</td>
</tr>
<tr>
<td>Symptomatic Medication</td>
<td>Only</td>
</tr>
</tbody>
</table>

Table A2

*MS Participant Group Information*

| How severe do you perceive your multiple sclerosis (MS) to be at present? | 58 | 1 | 5 | 5 | 3.88 | 1.08 |
| How severe do you perceive the condition of multiple sclerosis (MS) to be in general for most people? | 56 | 1 | 5 | 3 | 2.64 | 0.86 |
| Level of Functioning (Based on EDSS Score) | 58 | 0 | 8 | 0 | 3.319 | 2.75 |
### Skew and Kurtosis of the Data Set

**Table A3**

<table>
<thead>
<tr>
<th>Variable</th>
<th>$n$</th>
<th>Range</th>
<th>Min.</th>
<th>Max.</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Experiment 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Statistic</td>
<td>Std. Error</td>
</tr>
<tr>
<td>Health Incongruent</td>
<td>161</td>
<td>3</td>
<td>0.00</td>
<td>3.00</td>
<td>0.57</td>
<td>0.76</td>
<td>1.16</td>
<td>0.19</td>
</tr>
<tr>
<td>Health Congruent</td>
<td>161</td>
<td>4</td>
<td>0.00</td>
<td>4.00</td>
<td>0.89</td>
<td>0.85</td>
<td>0.76</td>
<td>0.19</td>
</tr>
<tr>
<td>MS Incongruent</td>
<td>161</td>
<td>4</td>
<td>0.00</td>
<td>4.00</td>
<td>0.54</td>
<td>0.84</td>
<td>1.67</td>
<td>0.19</td>
</tr>
<tr>
<td>MS Congruent</td>
<td>161</td>
<td>3</td>
<td>0.00</td>
<td>3.00</td>
<td>0.88</td>
<td>0.85</td>
<td>0.60</td>
<td>0.19</td>
</tr>
<tr>
<td>General Incongruent</td>
<td>161</td>
<td>4</td>
<td>0.00</td>
<td>4.00</td>
<td>0.45</td>
<td>0.72</td>
<td>1.96</td>
<td>0.19</td>
</tr>
<tr>
<td>General Congruent</td>
<td>161</td>
<td>3</td>
<td>0.00</td>
<td>3.00</td>
<td>0.34</td>
<td>0.57</td>
<td>1.91</td>
<td>0.19</td>
</tr>
</tbody>
</table>

=Assumption Violated
Table A3 cont.

**Skew and Kurtosis of the Data Set**

<table>
<thead>
<tr>
<th>Variable</th>
<th>$n$</th>
<th>Range</th>
<th>Min.</th>
<th>Max.</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Skewness Statistic</th>
<th>Std. Error</th>
<th>Kurtosis Statistic</th>
<th>Std. Error</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Experiment 2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MS Low Likelihood</td>
<td>161</td>
<td>1</td>
<td>0.00</td>
<td>1.00</td>
<td>0.32</td>
<td>0.47</td>
<td>0.76</td>
<td>0.19</td>
<td>-1.43</td>
<td>0.38</td>
</tr>
<tr>
<td>MS Med Likelihood</td>
<td>161</td>
<td>1</td>
<td>0.00</td>
<td>1.00</td>
<td>0.29</td>
<td>0.46</td>
<td>0.92</td>
<td>0.19</td>
<td>-1.16</td>
<td>0.38</td>
</tr>
<tr>
<td>MS High Likelihood</td>
<td>161</td>
<td>1</td>
<td>0.00</td>
<td>1.00</td>
<td>0.55</td>
<td>0.50</td>
<td>-0.19</td>
<td>0.19</td>
<td>-1.99</td>
<td>0.38</td>
</tr>
<tr>
<td>Health Low Likelihood</td>
<td>161</td>
<td>1</td>
<td>0.00</td>
<td>1.00</td>
<td>0.20</td>
<td>0.40</td>
<td>1.52</td>
<td>0.19</td>
<td>0.33</td>
<td>0.38</td>
</tr>
<tr>
<td>Health Med Likelihood</td>
<td>161</td>
<td>1</td>
<td>0.00</td>
<td>1.00</td>
<td>0.52</td>
<td>0.50</td>
<td>-0.09</td>
<td>0.19</td>
<td>-2.02</td>
<td>0.38</td>
</tr>
<tr>
<td>Health High Likelihood</td>
<td>161</td>
<td>1</td>
<td>0.00</td>
<td>1.00</td>
<td>0.30</td>
<td>0.46</td>
<td>0.86</td>
<td>0.19</td>
<td>-1.28</td>
<td>0.38</td>
</tr>
<tr>
<td>Social Low Likelihood</td>
<td>161</td>
<td>1</td>
<td>0.00</td>
<td>1.00</td>
<td>0.37</td>
<td>0.49</td>
<td>0.53</td>
<td>0.19</td>
<td>-1.74</td>
<td>0.38</td>
</tr>
<tr>
<td>Social Med Likelihood</td>
<td>161</td>
<td>1</td>
<td>0.00</td>
<td>1.00</td>
<td>0.57</td>
<td>0.50</td>
<td>-0.29</td>
<td>0.19</td>
<td>-1.94</td>
<td>0.38</td>
</tr>
<tr>
<td>Social High Likelihood</td>
<td>161</td>
<td>1</td>
<td>0.00</td>
<td>1.00</td>
<td>0.83</td>
<td>0.38</td>
<td>-1.74</td>
<td>0.19</td>
<td>1.03</td>
<td>0.38</td>
</tr>
<tr>
<td><strong>Experiment 3</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Error Number</td>
<td>149</td>
<td>130</td>
<td>-48.08</td>
<td>81.92</td>
<td>-20.85</td>
<td>26.14</td>
<td>1.46</td>
<td>0.20</td>
<td>1.52</td>
<td>0.39</td>
</tr>
</tbody>
</table>

=Assumption Violated
Appendix N

Moderator Analyses for Experiment 1

Table A4

*Pearson Correlations for Experiment 1*

<table>
<thead>
<tr>
<th></th>
<th>Health Congruent</th>
<th>Health Incongruent</th>
<th>MS Congruent</th>
<th>MS Incongruent</th>
<th>General Congruent</th>
<th>General Incongruent</th>
</tr>
</thead>
<tbody>
<tr>
<td>WHODAS Total Score</td>
<td>0.13</td>
<td>0.13</td>
<td>0.21**</td>
<td>0.13</td>
<td>0.18*</td>
<td>0.15</td>
</tr>
<tr>
<td>PHQ-9 Total Score</td>
<td>0.04</td>
<td>-0.02</td>
<td>0.01</td>
<td>-0.02</td>
<td>0.03</td>
<td>-0.007</td>
</tr>
<tr>
<td>Extroversion</td>
<td>0.09</td>
<td>0.06</td>
<td>0.02</td>
<td>0.07</td>
<td>0.06</td>
<td>0.07</td>
</tr>
<tr>
<td>Agreeableness</td>
<td>-0.07</td>
<td>0.02</td>
<td>-0.08</td>
<td>0.000</td>
<td>0.03</td>
<td>-0.004</td>
</tr>
<tr>
<td>Conscientiousness</td>
<td>-0.03</td>
<td>0.02</td>
<td>-0.008</td>
<td>-0.07</td>
<td>-0.01</td>
<td>0.03</td>
</tr>
<tr>
<td>Emotional Stability</td>
<td>0.08</td>
<td>0.15</td>
<td>0.11</td>
<td>0.16*</td>
<td>0.05</td>
<td>0.06</td>
</tr>
<tr>
<td>Openness</td>
<td>0.001</td>
<td>0.04</td>
<td>0.009</td>
<td>0.02</td>
<td>0.02</td>
<td>0.05</td>
</tr>
<tr>
<td>Currently in Education or Employment</td>
<td>-0.03</td>
<td>-0.09</td>
<td>-0.08</td>
<td>-0.08</td>
<td>-0.040</td>
<td>-0.06</td>
</tr>
</tbody>
</table>

**Correlation is significant at the 0.01 level; *0.05 level**
Table A5

*Pearson Correlations for Experiment 1 MS Group Variables*

<table>
<thead>
<tr>
<th></th>
<th>Health Congruent</th>
<th>Health Incongruent</th>
<th>MS Congruent</th>
<th>MS Incongruent</th>
<th>General Congruent</th>
<th>General Incongruent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Current Medication</td>
<td>Pearson</td>
<td>0.03</td>
<td>0.10</td>
<td>-0.02</td>
<td>0.09</td>
<td>-0.08</td>
</tr>
<tr>
<td></td>
<td>Sig.</td>
<td>.84</td>
<td>.463</td>
<td>.857</td>
<td>.493</td>
<td>.539</td>
</tr>
<tr>
<td>Severity of MS: Personal</td>
<td>Pearson</td>
<td>-0.21</td>
<td>-0.21</td>
<td>-.36**</td>
<td>-0.26</td>
<td>-0.05</td>
</tr>
<tr>
<td></td>
<td>Sig.</td>
<td>.114</td>
<td>.11</td>
<td>.005</td>
<td>.051</td>
<td>.712</td>
</tr>
<tr>
<td>Severity of MS: General</td>
<td>Pearson</td>
<td>0.06</td>
<td>-0.07</td>
<td>0.16</td>
<td>0.01</td>
<td>0.09</td>
</tr>
<tr>
<td></td>
<td>Sig.</td>
<td>.66</td>
<td>.586</td>
<td>.243</td>
<td>.935</td>
<td>.492</td>
</tr>
<tr>
<td>EDSS Score</td>
<td>Pearson</td>
<td>-0.06</td>
<td>0.07</td>
<td>0.26</td>
<td>0.14</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>Sig.</td>
<td>.647</td>
<td>.581</td>
<td>.053</td>
<td>.3</td>
<td>.687</td>
</tr>
</tbody>
</table>

**Correlation is significant at the 0.01 level; *0.05 level**
Appendix O

Covariate Analyses for Experiment 1

Table A6

Covariate Analysis for Experiment 1

<table>
<thead>
<tr>
<th>Source</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Emotional Stability</td>
<td>3.21</td>
<td>5.95</td>
<td>0.54</td>
<td>1.94</td>
<td>0.074</td>
</tr>
<tr>
<td>WHODAS Total Score</td>
<td>2.85</td>
<td>5.95</td>
<td>0.48</td>
<td>1.72</td>
<td>0.117</td>
</tr>
<tr>
<td>Level of Education</td>
<td>1.06</td>
<td>4.00</td>
<td>0.26</td>
<td>0.95</td>
<td>0.437</td>
</tr>
<tr>
<td>Error</td>
<td>86.36</td>
<td>309.62</td>
<td>0.28</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix P

Non-Parametric Test for Experiment 1

Table A7

*Mean Ranks for Friedman Test for Experiment 1*

<table>
<thead>
<tr>
<th>Risk Type</th>
<th>Congruence Format</th>
<th>Mean Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>MS</td>
<td>Congruent</td>
<td>4.10</td>
</tr>
<tr>
<td></td>
<td>Incongruent</td>
<td>3.28</td>
</tr>
<tr>
<td>Health</td>
<td>Congruent</td>
<td>4.09</td>
</tr>
<tr>
<td></td>
<td>Incongruent</td>
<td>3.42</td>
</tr>
<tr>
<td>Social</td>
<td>Congruent</td>
<td>2.95</td>
</tr>
<tr>
<td></td>
<td>Incongruent</td>
<td>3.16</td>
</tr>
</tbody>
</table>

Table A8

*Friedman Test Results for Experiment 1*

<table>
<thead>
<tr>
<th>n</th>
<th>Chi-Square</th>
<th>df</th>
<th>Asymp. Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>161</td>
<td>89.94</td>
<td>5</td>
<td>0.00</td>
</tr>
</tbody>
</table>
Appendix Q

Moderator Analyses for Experiment 2

Table A9

Pearson Correlations for Experiment 2

<table>
<thead>
<tr>
<th></th>
<th>MS Low</th>
<th>MS Medium</th>
<th>MS High</th>
<th>Health Low</th>
<th>Health Medium</th>
<th>Health High</th>
<th>Social Low</th>
<th>Social Medium</th>
<th>Social High</th>
</tr>
</thead>
<tbody>
<tr>
<td>WHODAS Total Score</td>
<td>.24**</td>
<td>0.08</td>
<td>0.06</td>
<td>0.10</td>
<td>0.09</td>
<td>.17*</td>
<td>0.07</td>
<td>0.03</td>
<td>-0.11</td>
</tr>
<tr>
<td>PHQ-9 Total Score</td>
<td>-0.001</td>
<td>0.06</td>
<td>0.12</td>
<td>0.02</td>
<td>0.04</td>
<td>0.05</td>
<td>0.06</td>
<td>-0.04</td>
<td>-0.03</td>
</tr>
<tr>
<td>Extroversion</td>
<td>0.11</td>
<td>0.02</td>
<td>-0.04</td>
<td>0.09</td>
<td>-0.09</td>
<td>0.03</td>
<td>0.02</td>
<td>0.003</td>
<td>0.06</td>
</tr>
<tr>
<td>Agreeableness</td>
<td>0.12</td>
<td>0.10</td>
<td>-0.01</td>
<td>0.01</td>
<td>-0.03</td>
<td>0.07</td>
<td>-0.02</td>
<td>-0.02</td>
<td>-0.04</td>
</tr>
<tr>
<td>Conscientiousness</td>
<td>.21**</td>
<td>0.04</td>
<td>-0.07</td>
<td>-0.02</td>
<td>-0.09</td>
<td>0.05</td>
<td>-0.05</td>
<td>-0.13</td>
<td>-0.08</td>
</tr>
<tr>
<td>Emotional Stability</td>
<td>.23**</td>
<td>0.03</td>
<td>-0.05</td>
<td>0.09</td>
<td>-0.13</td>
<td>0.10</td>
<td>-0.16*</td>
<td>-0.03</td>
<td>-0.11</td>
</tr>
<tr>
<td>Openness</td>
<td>0.13</td>
<td>-0.05</td>
<td>-0.07</td>
<td>0.003</td>
<td>-0.08</td>
<td>-0.04</td>
<td>-0.08</td>
<td>-0.08</td>
<td>-0.09</td>
</tr>
<tr>
<td>Highest Level of Education</td>
<td>-0.15</td>
<td>-0.09</td>
<td>0.09</td>
<td>-0.15</td>
<td>0.10</td>
<td>-0.10</td>
<td>0.07</td>
<td>0.09</td>
<td>0.12</td>
</tr>
<tr>
<td>Currently in Education or Employment</td>
<td>-.19*</td>
<td>-0.07</td>
<td>0.11</td>
<td>-0.12</td>
<td>-0.07</td>
<td>-0.02</td>
<td>-.20*</td>
<td>0.11</td>
<td>0.04</td>
</tr>
</tbody>
</table>

**Correlation is significant at the 0.01 level; *0.05 level
Table A10

*Pearson Correlations for Experiment 2 MS Group Variables*

<table>
<thead>
<tr>
<th></th>
<th>MS</th>
<th>MS</th>
<th>MS</th>
<th>Health</th>
<th>Health</th>
<th>Health</th>
<th>Social</th>
<th>Social</th>
<th>Social</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low</td>
<td>Medium</td>
<td>High</td>
<td>Low</td>
<td>Medium</td>
<td>High</td>
<td>Low</td>
<td>Medium</td>
<td>High</td>
</tr>
<tr>
<td>Current Medication</td>
<td>Pearson</td>
<td>0.17</td>
<td>-0.17</td>
<td>-0.16</td>
<td>-0.2</td>
<td>0.10</td>
<td>-0.16</td>
<td>0.001</td>
<td>-0.002</td>
</tr>
<tr>
<td></td>
<td>Sig.</td>
<td>.193</td>
<td>.196</td>
<td>.23</td>
<td>.133</td>
<td>.47</td>
<td>.219</td>
<td>.996</td>
<td>.986</td>
</tr>
<tr>
<td>Severity of MS: Personal</td>
<td>Pearson</td>
<td>0.008</td>
<td>0.008</td>
<td>0.02</td>
<td>-0.08</td>
<td>0.02</td>
<td>-0.18</td>
<td>-0.02</td>
<td>-0.03</td>
</tr>
<tr>
<td></td>
<td>Sig.</td>
<td>.95</td>
<td>.95</td>
<td>.907</td>
<td>.577</td>
<td>.881</td>
<td>.182</td>
<td>.858</td>
<td>.849</td>
</tr>
<tr>
<td>Severity of MS: General</td>
<td>Pearson</td>
<td>-0.16</td>
<td>-0.07</td>
<td>-0.11</td>
<td>-0.06</td>
<td>0.14</td>
<td>-0.11</td>
<td>-0.003</td>
<td>0.14</td>
</tr>
<tr>
<td></td>
<td>Sig.</td>
<td>.252</td>
<td>.599</td>
<td>.428</td>
<td>.663</td>
<td>.302</td>
<td>.428</td>
<td>.982</td>
<td>.323</td>
</tr>
<tr>
<td>EDSS Score</td>
<td>Pearson</td>
<td>-0.03</td>
<td>0.12</td>
<td>0.08</td>
<td>0.18</td>
<td>-0.12</td>
<td>.28*</td>
<td>0.15</td>
<td>-0.16</td>
</tr>
<tr>
<td></td>
<td>Sig.</td>
<td>.838</td>
<td>.379</td>
<td>.554</td>
<td>.168</td>
<td>.369</td>
<td>.036</td>
<td>.272</td>
<td>.235</td>
</tr>
</tbody>
</table>

**Correlation is significant at the 0.01 level; *0.05 level**
### Appendix R

**Covariate Analyses for Experiment 2**

**Table A11**

*Covariate Analyses for Experiment 2*

<table>
<thead>
<tr>
<th>Source</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig</th>
</tr>
</thead>
<tbody>
<tr>
<td>Emotional Stability</td>
<td>Greenhouse-Geisser</td>
<td>0.66</td>
<td>3.25</td>
<td>0.20</td>
<td>0.87</td>
</tr>
<tr>
<td>WHODAS Total Score</td>
<td>Greenhouse-Geisser</td>
<td>0.41</td>
<td>3.25</td>
<td>0.13</td>
<td>0.54</td>
</tr>
<tr>
<td>Currently in Education or Employment</td>
<td>Greenhouse-Geisser</td>
<td>1.79</td>
<td>3.25</td>
<td>0.55</td>
<td>2.36</td>
</tr>
<tr>
<td>Error</td>
<td>Greenhouse-Geisser</td>
<td>111.35</td>
<td>477.35</td>
<td>0.23</td>
<td></td>
</tr>
</tbody>
</table>
Appendix S

Non-Parametric Test for Experiment 2

Table A12

*Mean Ranks for Friedman Test for Experiment 2*

<table>
<thead>
<tr>
<th>Risk Type</th>
<th>Risk Level</th>
<th>Mean Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>MS</td>
<td>Low Likelihood</td>
<td>4.48</td>
</tr>
<tr>
<td></td>
<td>Medium Likelihood</td>
<td>4.34</td>
</tr>
<tr>
<td></td>
<td>High Likelihood</td>
<td>5.48</td>
</tr>
<tr>
<td>Health</td>
<td>Low Likelihood</td>
<td>3.92</td>
</tr>
<tr>
<td></td>
<td>Medium Likelihood</td>
<td>5.37</td>
</tr>
<tr>
<td></td>
<td>High Likelihood</td>
<td>4.39</td>
</tr>
<tr>
<td>Social</td>
<td>Low Likelihood</td>
<td>4.70</td>
</tr>
<tr>
<td></td>
<td>Medium Likelihood</td>
<td>5.59</td>
</tr>
<tr>
<td></td>
<td>High Likelihood</td>
<td>6.74</td>
</tr>
</tbody>
</table>

Table A13

*Friedman Test Results for Experiment 2*

<table>
<thead>
<tr>
<th>n</th>
<th>Chi-Square</th>
<th>df</th>
<th>Asymp. Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>161</td>
<td>220.22</td>
<td>8</td>
<td>0.00</td>
</tr>
</tbody>
</table>
Appendix T

NHS Research Ethics Committee: Favourable Ethical Opinion Letter

(Document Redacted)
Appendix U

University Hospitals Birmingham Research Governance: Project Authorisation Letter

(Document Redacted)