`

An Associative Account of Impaired Inhibition in Psychopathy

Development of the Psychopathy Attention Theory

Georgina Mellody Aisbitt

Brasenose College, University of Oxford

Hilary 2016



Thesis submitted in support of a DPhil in Experimental Psychology,

University of Oxford

ACKNOWLEDGEMENTS

I wish to thank everyone who has encouraged, inspired and challenged me.

My sincerest thanks to Robin Murphy, my supervisor, without whose encouragement I would not have considered pursuing this DPhil. I wish to thank him for challenging me, teaching me to question what I know and helping me not to fear either process. I thank him for giving me the freedom to think and to learn and for his resolute patience in both (even when resorting to aardvark cartoons!). Robin's guidance and knowledge has been invaluable, and I am extremely thankful to have had him as my supervisor.

I wish to thank David Popplewell, my college advisor, for his support throughout both my undergraduate degree and this DPhil. Dave has been a huge source of encouragement, particularly whilst I made the transition to postgraduate study, and I am grateful to him for involving me in the college life and for 'having my back' in research, welfare and even on the river. Thank you to Sadahiko Nakajima for acting as my supervisor during my Summer Research Fellowship at Kwansei Gakuin University. Nakajima-sensei went above and beyond to assist me in achieving my research aims, but also in ensuring that I enjoyed my time in Japan. I am very appreciative of how welcome he made me feel, and the opportunity I had to discuss my research with him.

I also wish to thank Rachel Msetfi, Nicola Byrom and Dyedra Just for their advice and reassurance throughout this process. They have each been generous with their time, and I am sincerely grateful for that. Thank you to Brasenose College Annual Fund, the

John Fell Fund, the Japan Society for the Promotion of Science and the Experimental Psychology Society for their generous financial assistance with this project.

Thank you to Jon and Julia Aisbitt, my wonderful parents, for their unwavering support during this DPhil. and the many years preceding it. I thank them for always encouraging me and for showing me the beauty of learning. Thank you to my father for his esteemed guidance and for always agreeing to proofread my writing; and, thank you to my mother for always asking how I am and really listening to answer (even if it dragged on!). Together they have inspired me to embrace the challenge of the unknown for the possibility it holds, and I am sincerely grateful for that. I also wish to thank Lexi and Toby Aisbitt, my siblings, for their loving support. I thank them for always making time for me (whatever the time difference) and for encouraging me to make time for myself. I am so very grateful for my family and all that they have done for me.

Thank you to Callum McPherson, my partner, for his untiring support day in, day out, whether piloting new programmes or proofreading this thesis. I thank Callum for showing a genuine interest in what I do, and for his sincere efforts to understand and engage with my research. I am grateful to Callum for his kindness, love and encouragement during this journey, and for never asking more than I have been able to give.

CONTENTS

PAGE

TITLE PAGE ..... **i**

ACKNOWLEDGMENTS ..... **ii**

ABSTRACT ..... **ix**

LIST OF TABLES ..... **x**

LIST OF FIGURES ..... **xi**

OVERVIEW ..... **xiii**

CHAPTER 1 – INTRODUCTION **15**

1.1. Psychopathic Personality **15**

1.2. Reversal Learning Differences in Psychopathy **17**

1.3. Theories of Psychopathy **20**

1.3.1. Outcome Processing Theories **20**

1.3.2. Attentional Theories **22**

1.4. Impaired Integration Theory **24**

1.5. Learning Account of Psychopathy **27**

1.6. Explaining Reversal Inconsistencies **33**

1.7. Assessing Inhibition **35**

1.8. Novel Method for Assessing Learning in Psychopathy **39**

CHAPTER 2 – ASSESSING LEARNING IN PSYCHOPATHY (EXPERIMENT 1) **42**

2.1. Introduction **42**

2.2. Method **45**

2.2.1. Participants **45**

2.2.2. Design **46**

2.2.3. Materials **46**

2.2.4. Procedure **51**

2.2.5. Statistical Analysis **51**

2.3. Results **53**

2.3.1. Overall Learning Effects **53**

2.3.2. Relationship Between Psychopathy and Dependent Variables **56**

2.4. Discussion **57**

2.5. Conclusions **58**

CHAPTER 3 – TESTING THE CONDITIONS OF LEARNING DIFFERENCES IN PSYCHOPATHY (EXPERIMENTS 2-5) **59**

3.1. Introduction **59**

3.2. General Method **60**

3.2.1. Psychopathic Personality Inventory **60**

3.2.2. Novel Shift Go/NoGo Task **61**

3.2.3. Statistical Analysis **64**

3.3. Experiment 2 ..... **68**

3.3.1. Introduction **68**

3.3.2. Method **69**

3.3.2.1. Participants **69**

3.3.2.2. Design **70**

3.3.2.3. Materials **70**

3.3.2.4. Procedure **73**

3.3.3. Results **73**

3.3.3.1. Overall Learning Effects **74**

3.3.3.2. Relationship Between Psychopathy and Dependent Variables **77**

3.3.4. Discussion **79**

3.4. Experiment 3 ..... **82**

3.4.1. Introduction **82**

3.4.2. Method **82**

3.4.2.1. Participants and Design **82**

3.4.2.2. Materials **83**

3.4.2.3. Procedure **85**

3.4.3. Results **85**

3.4.3.1. Overall Learning Effects **86**

3.4.3.2. Relationship Between Psychopathy and Dependent Variables **90**

3.4.3.3. *Post Hoc* Analysis of the Inhibitory Cue **93**

3.4.4. Discussion **95**

3.5. Experiment 4 ..... **100**

3.5.1. Introduction **100**

3.5.2. Method **101**

3.5.2.1. Participants and Design **101**

3.5.2.2. Materials **101**

3.5.3. Results **103**

3.5.3.1. Overall Learning Effects **103**

3.5.3.2. Relationship Between Psychopathy and Dependent Variables **107**

3.5.3.3. *Post Hoc* Analysis of the Inhibitory Cue **108**

3.5.4. Discussion **109**

3.6. Experiment 5 ..... **114**

3.6.1. Introduction **114**

3.6.2. Method **115**

3.6.2.1. Participants **115**

3.6.2.2. Design **115**

3.6.2.3. Materials **115**

3.6.3. Results **118**

3.6.3.1. Overall Learning Effects **118**

3.6.3.2. Relationship Between Psychopathy and Dependent Variables **123**

3.6.3.3. *Post Hoc* Analysis of the Inhibitory Cue **125**

3.6.4. Discussion **126**

3.7. Summary of Chapter 3 Results **128**

3.8. General Discussion **131**

3.9. Conclusions **134**

CHAPTER 4 – THE EFFECT OF COGNITIVE LOAD ON SHIFT LEARNING (EXPERIMENT 6) **136**

4.1. Introduction **136**

4.2. Method **138**

4.2.1. Participants and Design **138**

4.2.2. Materials **139**

4.2.3. Statistical Analysis **139**

4.3. Results **140**

4.3.1. Overall Learning Effects **141**

4.3.2. Relationship Between Psychopathy and Dependent Variables **143**

4.3.3. Post-Hoc Comparison of Experiments 3 and 6 ..... **144**

4.4. Discussion **145**

4.5. Conclusions **147**

CHAPTER 5 – CONTEXTUAL PROCESSING AND PERCEPTIONS OF CONTROL (EXPERIMENT 7) **148**

5.1. Introduction **148**

5.2. Method **152**

5.2.1. Participants **152**

5.2.2. Design **152**

5.2.3. Materials **153**

5.2.4. Procedure **156**

5.2.5. Statistical Analysis **157**

5.3. Results **159**

5.3.1. Overall Learning Effects **159**

5.3.2. Relationship Between Psychopathy and Dependent Variables **162**

5.4. Discussion **164**

5.5. Conclusions **168**

CHAPTER 6 – GENERAL DISCUSSION **169**

6.1. Impaired Integration Theory **171**

6.2. Theory of Attention **172**

6.3. Novel Proposal: The Psychopathy Attention Theory **174**

6.4. Future Experiments to Test the Psychopathy Attention Theory **180**

6.5. Conceptualisation of Psychopathy **182**

6.6. Assessing Inhibition **185**

6.7. Negative Super-Learning **187**

6.8. Conclusions **189**

REFERENCES **191**

ABSTRACT

An Associative Account of Impaired Inhibition in Psychopathy:

Development of the Psychopathy Attention Theory

Georgina Mellody Aisbitt

Brasenose College, University of Oxford

Hilary 2016

Thesis submitted in support of a D.Phil. in Experimental Psychology

Psychopathy, a multi-faceted construct characterised by antisocial behaviours and callous-unemotional traits, is associated with atypical learning. However, the nature of this atypical learning is unclear due to inconsistencies in the literature. Attentional and outcome processing theories of psychopathy do not provide an adequate explanation of these inconsistencies, but I argue that an attentional-associative account can. For instance, Mackintosh's (1975) theory can be used to interpret learning differences in psychopathy through disruption in establishing a cue's associability: a representation of the relative validity of a cue used to determine attentional allocation, and so how easily the cue is learnt about. The discrepancy between previous studies can be attributed to methodological differences and how these differentially influence the associability of a stimulus. In a series of experiments I test this hypothesis and the conditions under which learning differences are seen in psychopathy. I report that psychopathic traits predict altered learning about inhibitory cues, but not excitatory cues, and show that these effects are sensitive to manipulation of predictiveness and cognitive load. By incorporating principles of reduced processing capacity in psychopathic individuals into an attentional-associative model, I am able to explain the results of my experiments, and provide an account for previous discrepancies in the literature.

LIST OF TABLES

PAGE

Table 1. Experiment 1: Participant Psychopathy *T*-Scores **46**

Table 2. Experiment 1: Design **48**

Table 3. Experiment 1: Zero-Order Correlations **56**

Table 4. Experiment 2: Participant Psychopathy *T*-Scores **70**

Table 5. Experiment 2: Design **71**

Table 6. Experiment 2: Zero-Order Correlations **78**

Table 7. Experiment 3: Participant Psychopathy *T*-Scores **83**

Table 8. Experiment 3: Design **84**

Table 9. Experiment 3: Zero-Order Correlations **90**

Table 10. Experiment 4: Participant Psychopathy *T*-Scores **101**

Table 11. Experiment 4: Design **102**

Table 12. Experiment 4: Zero-Order Correlations **107**

Table 13. Experiment 5: Participant Psychopathy *T*-Scores **116**

Table 14. Experiment 5: Design **117**

Table 15. Experiment 5: Zero-Order Correlations **123**

Table 16. Experiment 6: Participant Psychopathy *T*-Scores **139**

Table 17. Experiment 6: Zero-Order Correlations **144**

Table 18. Experiment 7: Participant Psychopathy *T*-Scores **153**

Table 19. Experiment 7: Zero-Order Correlations **162**

LIST OF FIGURES

PAGE

Figure 1. Mediation Model of Hypothesised Effect Between Psychopathy and Super-Learning **45**

Figure 2. Experiment 1: Example Trial Sequence **50**

Figure 3. Experiment 1: Initial Cue Learning **54**

Figure 4. Experiment 1: Sex Differences in Positive Super-Learning **55**

Figure 5. Experiments 2-6: Example Trial Sequence **61**

Figure 6. Experiment 2: Inhibitory Learning Effects **75**

Figure 7. Experiment 2: Excitatory and Inhibitory Shift Learning **76**

Figure 8. Experiment 3: Initial Cue Learning **86**

Figure 9. Experiment 3: Excitatory and Inhibitory Shift Learning **87**

Figure 10. Experiment 3: Positive and Negative Super-Learning **89**

Figure 11. Experiment 3: Mediation Analysis of Fearless Dominance **92**

Figure 12. Experiment 4: Initial Cue Learning **103**

Figure 13. Experiment 4: Excitatory and Inhibitory Shift Learning **105**

Figure 14. Experiment 4: Positive and Negative Super-Learning **106**

Figure 15. Experiment 5: Initial Cue Learning **119**

Figure 16. Experiment 5: Excitatory, Inhibitory and Neutral Shift Learning **120**

Figure 17. Experiment 5: Positive and Negative Super-Learning **122**

Figure 18. Experiment 5: Mediation Analysis of Fearless Dominance **125**

Figure 19: Experiment 6: Processing Bottleneck Diagrams **137**

Figure 20. Experiment 6: Initial Cue Learning **141**

Figure 21. Experiment 6: Excitatory and Inhibitory Shift Learning **142**

Figure 22. Experiment 7: Positive and Zero Contingencies **154**

Figure 23. Experiment 7: Attribution of Control to the Context or One's Actions **161**

Figure 24. Experiment 7: Mediation Analysis of Fearless Dominance **163**

OVERVIEW

Chapter 1 of this thesis will begin by presenting a description of the traits and behaviours that characterise psychopathic personality. The focus will be on reversal learning, one of the well-documented features of psychopathic impulsivity. The role of reversal learning tests in understanding psychopathy will be discussed, and the results interpreted through existing outcome processing and attentional accounts of psychopathy. Having identified limitations in these accounts, reversal learning will be interpreted in light of the recently proposed Impaired Integration Theory (Hamilton, Hiatt Racer, & Newman, 2015) of psychopathy.

Following this, Chapter 1 will continue with a case being made for assessing reversal learning in psychopathy through its relationship to principles of attention and associative learning, such as those outlined in Mackintosh's (1975) Theory of Attention. I will give a brief overview of Mackintosh's theory, its key principles and examples of supporting evidence from the animal and human literature. I will then interpret the altered psychopathy reversal data using this alternative model. I will discuss the way in which methodological variations might affect the principles of Mackintosh's theory, and so provide a possible account of the inconsistency of previous findings.

Chapters 2-5 will present evidence from a series of experiments designed to test the associative and attentional features that affect shift learning in relation to psychopathic traits (Aisbitt & Murphy, submitted a, Chapter 3; Aisbitt, Msetfi & Murphy, submitted, Chapter 5). In Chapter 6, I will argue that a theory of attention and learning can incorporate the principles of the Impaired Integration Theory (Hamilton et al., 2015), particularly the idea of an information-processing bottleneck, in order to explain the

present data and to mechanise the concepts described in the Impaired Integration Theory. A novel Psychopathy Attention Theory (Aisbitt & Murphy, in press. b, Chapters 3 and 6) designed to incorporate the principles of the Impaired Integration Theory into an attentional-associative account will be proposed. I will argue that this novel framework can be used to explain previous inconsistencies in the literature. The limitations and implications of these results will be discussed, and suggestions for future research will be made.

CHAPTER 1

Introduction

1.1. Psychopathic Personality

Psychopathic personality, or psychopathy, is generally defined by the combination of impulsive-antisocial behaviours (e.g., promiscuous sexual behaviour, juvenile delinquency, parasitic lifestyle) and callous-unemotional traits (e.g., pathological lying, shallow affect; Cleckley, 1982; Hare & Neumann, 2008; Hare, 1991; Lilienfeld & Widows, 2005). Psychopathy appears in the DSM-V (American Psychiatric Association, 2013) under the Antisocial Personality Disorders section. However, the conceptualisation of psychopathy in the DSM has been criticised for focusing too much on antisocial and delinquent behaviours, and not enough on the stable personality traits also believed to define psychopathy (Hare, Hart, & Harpur, 1991).

Hare's Psychopathic Checklist (Revised; PCL-R; Hare, 1991) includes four such personality dimensions: Interpersonal (grandiosity, glibness, deceitfulness, manipulative); Affective (blame externalisation, callous/lack of empathy, shallow affect); Lifestyle (impulsive, irresponsible, lack of realistic long-term goals, parasitic lifestyle); and, Antisocial (antisocial behaviour, poor behavioural control). These four dimensions load on to two factors: Factor 1 represents the Interpersonal/Affective dimensions; and, Factor 2 represents the Lifestyle/Antisocial dimensions (Hare & Neumann, 2008). Whilst both factors are strongly related to some types of behaviours, such as violent recidivism

(Hemphill, Hare, & Wong, 1998), the factors are differentially related to other core cognitive and behavioural processes. For instance, Factor 1 is linked to social dominance (Verona, Patrick, & Joiner, 2001), whilst Factor 2 is predictive of 'spontaneous and disinhibited violence' (Hart & Dempster, 1997) and destructive impulsive behaviour (Harpur, Hare, & Hakstian, 1989).

However, this two-factor structure derived from the PCL-R has been criticised for failing to capture other positive adjustment features of psychopathy (Patrick, Fowles, & Krueger, 2009). For example, psychopathic individuals are better able than non-psychopathic controls to shift sorting strategies following rule changes in the Wisconsin Card Sorting Task, an ability that is suggestive of enhanced executive functioning (Ishikawa, Raine, Lencz, Bihrle, & Lacasse, 2001). Although this improved executive functioning has also been linked to the capacity to lie to and manipulate others (Gao & Raine, 2010), it still represents enhanced performance compared to non-psychopathic controls.

An assessment tool that does capture both negative and positive psychopathy features is the Psychopathic Personality Inventory (Revised; PPI-R; Lilienfeld & Widows, 2005). Overall PPI-R scores are grouped into three factors: Fearless Dominance; Self-Centred Impulsivity; and, Coldheartedness. Fearless Dominance (FD) reflects positive psychological and social adjustments, such as charm, as well as capturing negative adjustment tendencies, such as fearlessness (Benning, Patrick, Hicks, Blonigen, & Krueger, 2003; Patrick et al., 2009), evidenced by the failure to show appropriate fear conditioning (e.g., López, Poy, Patrick, & Moltó, 2013). Self-Centred Impulsivity (SCI) reflects psychological and behavioural maladjustment, including impulsivity, shown, for instance, through increased responses to cues signalling a need to withhold a response (e.g., Budhani, Richell, & Blair, 2006), even when this results in the

loss of a reward (Newman, Patterson, Howland, & Nichols, 1990; Newman, Patterson, & Kosson, 1987). The final factor of Coldheartedness (C) relates exclusively to the absence of guilt and empathy and is not well correlated with the other two factors (S. O. Lilienfeld & Widows, 2005).

The three PPI-R factors broadly map on to those identified by Patrick et al.'s (2009) triarchic model, which refers to the factors as: Boldness; Disinhibition; and, Meanness respectively. Like the PPI-R or Patrick et al.'s triarchic model, other three-factor models, such as that proposed by Cooke & Michie (2001) which is based on the PCL-R, also include a factor that is particularly associated with impulsivity and disinhibition. As such, it is clear that conceptualisations of psychopathy consider impulsivity to be a central component, and indeed it is one of its most studied characteristics, relating generally to both impulsivity (e.g., R. J. R. Blair et al., 2004; Newman & Schmitt, 1998), and more specifically to poor inhibition (e.g., He, Cassaday, Howard, Khalifa, & Bonardi, 2011).

1.2. Reversal Learning Differences in Psychopathy

Learning about the consequences of actions, and adapting behaviour on the basis of this learning, is essential for enabling adaptive interaction with the environment (e.g., Byrom, Msetfi, & Murphy, 2015). This learning involves forming associations between stimuli (S), responses (R) and outcomes (O), so that participants learn how to respond to specific stimuli in order to achieve a desired outcome. For example, people can learn to respond to some stimuli and withhold responding to other stimuli in order to achieve reward and/or avoid punishment (Chatlosh, Neunaber, & Wasserman, 1985). More than the ability to acquire these S-R-O associations, people are able to flexibly adjust these associations and appear sensitive to changes in the contingency between responding and

its consequences for a given stimulus (Clark, Cools, & Robbins, 2004), even when there is some variability about the contingency between stimuli, responses and outcomes (e.g., probabilistic reinforcement; Cools, Clark, Owen, & Robbins, 2002). For instance, in order to avoid punishment, people learn to withhold a response to a stimulus that previously signalled the need to make a response (Newman et al., 1987).

However, there are considerable individual differences in this ability to acquire and adjust S-R-O associations, and the apparent disruption of this learning is considered a defining feature of psychopathy, and more specifically psychopathic impulsivity (Cleckley, 1982; Hare, 1991; S. O. Lilienfeld & Widows, 2005; Patrick et al., 2009). For instance, the view that psychopathic individuals are impulsive is supported by evidence of their disrupted ability to learn when to respond and when to withhold responding to stimuli in order to achieve reward and/or avoid punishment, abilities commonly tested using a go/nogo task (Newman & Kosson, 1986; Newman, Widom, & Nathan, 1985). For example, in a task requiring participants to respond to squares, but to withhold their responding to crosses, psychopathic participants make significantly more incorrect responses to the crosses than their non-psychopathic comparisons do (Lapierre, Braun, & Hodgins, 1995). Evidence such as this, that psychopathic individuals: make more errors of commission, incorrect responses to 'nogo' cues (e.g., crosses) signalling the need to withhold the response; but, equivalent errors of omission, the incorrect withholding of a response to a 'go' cue (e.g., square) signalling the need to respond, has been regarded as a demonstration of impulsivity (Newman & Kosson, 1986).

However, the view that psychopaths are impulsive has been criticised for being overly simplistic and failing to recognise the complex relationship between psychopathy and impulsivity (Poythress & Hall, 2011). The relationship is considered complex because evidence of impulsivity, such as that outlined above, is inconsistent. For

instance: some studies report that psychopathy is associated with difficulties in learning when to respond and when to withhold responding (Newman & Kosson, 1986); yet, others report no difficulty in forming these initial associations, but do report psychopathy differences in how these associations are flexibly shifted in response to S-R-O contingency changes (e.g., Budhani et al., 2006; Mitchell, Colledge, Leonard, & Blair, 2002).

One study reporting such reversal-specific differences compared the performance of forensic psychopathic and non-psychopathic samples on a probabilistic object discrimination task (Budhani et al., 2006). Participants were exposed to both: a deterministic contingency, in which a response to a given cue always resulted in the same outcome, for instance a response was followed by reward on 100% of trials, and was followed by punishment on 0% of trials (i.e., 100-0 contingency); and, a probabilistic contingency, in which overall a response to a given cue resulted in an outcome but there was a degree of variability about this, for instance a response was followed by reward on 80% of trials, and punishment on 20% of the trials (i.e., 80-20 contingency). Both types of reinforcement schedule can be used to train S-R-O contingencies, and were used to signal which of two simultaneously presented stimuli participants should respond to.

Following this initial acquisition phase, these contingencies were reversed, and Budhani et al. (2006) measured participants' ability to monitor these contingency changes and adjust their responding to the new contingencies. Budhani and colleagues reported that although the psychopathy group and control group did not significantly differ in the acquisition phase, in the reversal phase the psychopathy group made significantly more errors than the control group. This was seen as evidence of psychopathic individuals' poor ability to monitor contingency changes and use this

information to guide their responding (e.g., Blair, Peschardt, Budhani, Mitchell, & Pine, 2006).

However, reversal differences such as this are not consistently observed, and some studies report no differences in the learning and behaviour of high psychopathy samples in either the acquisition or the reversal of contingencies (Gregory et al., 2015; Kiehl, Smith, Hare, & Liddle, 2000; Sadeh & Verona, 2008). For example, Gregory and colleagues compared the performance of psychopathic and non-psychopathic forensic samples on a task closely mirroring that used by Budhani and colleagues (2006) in the previous example, except that it was conducted in an MRI scanner. Once again, participants completed a probabilistic response reversal task, in which they received the same deterministic contingency (100-0) and probabilistic contingency (80-20) conditions. But, in contrast to the results of Budhani et al., Gregory et al. reported that the errors and response latencies of the psychopathy group were comparable to those of the control group in both the acquisition phase and the reversal phase. These results suggest that psychopathic individuals do not struggle to monitor changes in S-R-O contingencies. As such, it can be seen that the extent of learning and reversal differences in psychopathy are inconsistently reported between studies, and sometimes even within them (e.g., Brazil, Maes, et al., 2013).

1.3. Theories of Psychopathy

1.3.1. Outcome Processing Theories

One group of theories claiming to account for reversal-specific differences are referred to as 'outcome processing' or emotional-cognitive accounts of psychopathy. Two examples of such accounts, The Integrated Emotion Systems Model (IES; Blair, 2005) and the Differential Amygdala Activation Hypothesis (DAAM; Moul, Killcross, &

Dadds, 2012), both suggest that the ability to associate an outcome (O) with either a stimulus (S) or a response (R) is disrupted in psychopathic individuals. However, the two theories differ in their explanations of why this disruption occurs and how they account for the reversal specificity of the difference. The IES suggests that disruptions in the amygdala impair psychopathic individuals' ability to form S-O associations, but not their ability to form S-R associations (R. J. R. Blair, 2005). Blair suggests that this amygdala disruption uniquely affects psychopathic individuals' ability to encode the affective valence of outcomes, meaning that goal-directed learning, which requires sensitivity to the value of an outcome, cannot occur (Dickinson, 1985). In contrast, psychopathic individuals' habit behaviour, which is dependent on S-R associations, is intact.

It is this distinction between habit-like and goal-directed abilities that, Blair (2005) argues, can account for the reversal specific differences in psychopathy: initial learning in the acquisition phase is considered to depend only on habit-like learning; but, monitoring changes in contingencies in the reversal phase requires goal-directed learning and the processing of the outcome. However, whilst learning in the acquisition phase may rapidly shift to become habit-like in nature, it will initially depend on participants' ability to process the outcome in order to establish which response is correct. As such, Blair's account cannot provide an adequate explanation, unless it is assumed that some degree of goal-directed learning is preserved, but that this is not sufficient to overcome established habit-like responses.

The IES (R. J. R. Blair, 2005) can account for the discrepancy in the findings of Budhani et al. (2006) and Gregory et al. (2015) by examining the number of trials that are used in the two studies. Budhani and colleagues presented participants with 40 trials of each stimulus in each phase, whilst Gregory and colleagues presented participants with only 20 trials of the deterministic contingency, and 25 of the probabilistic

contingency in each phase. I have suggested that the IES requires a modification, such that goal-directed learning is not completely impaired, but is relatively unable to supersede habit-like processes. Following on from this, it could be argued that the habits formed after 20-25 trials (as in Gregory et al.'s study) will be substantially weaker than those formed after 40 trials (as in Budhani et al.'s study), and so the impoverished goal-directed process will be more likely to overcome the habit-like responses when these habits are weaker. As such, psychopathy differences are seen in Budhani et al.'s study, but not Gregory et al.'s study.

Another outcome processing theory, the DAAM, can also account for reversal differences in psychopathy (Moul et al., 2012). Moul and colleagues suggest that amygdala disruptions mean that generalised valence associations govern learning in psychopathy, and that the specific features of outcomes are ignored. As a result, whilst psychopathic individuals can distinguish valence (punishment vs. reward), they cannot distinguish value (degrees of punishment and reward). Therefore, they are able to acquire initial associations, but are insensitive to subsequent changes in the value of the outcome, and so do not update their learning when the S-R-O contingencies change. An adequate explanation for the variability in reversal learning may be developed along the lines of that offered in relation to the IES: that the strength of the associations formed in Budhani et al.'s (2006) study were far greater than those formed in Gregory et al.'s (2015) study, which made them less amenable to change. However, like the IES, the DAAM cannot account well for the differences in reversal learning, and this variability is not a prediction that is made by the model, and is instead accounted for here *post hoc*.

1.3.2. Attentional Theories

An alternative type of theory, that is perhaps better able to cope with the variability in the reported learning differences, suggests that psychopathic individuals have overly

selective attention, and that variations in a task's attentional demands determines performance (e.g., Response Modulation Hypothesis; Gorenstein & Newman, 1980; Attentional Bottleneck Model; Newman & Baskin-Sommers, 2011). Attentional accounts, such as the Response Modulation Hypothesis (RMH; Gorenstein & Newman, 1980), suggest that psychopathy is characterised by a relative processing deficiency or blindness for cues outside the focus of attention. These differences are particularly pronounced when the peripheral information is in conflict with the pursuit of a current goal (MacCoon, Wallace, & Newman, 2004; Newman, 1998).

Therefore, like the aforementioned outcome processing accounts, these attentional theories suggest that impulsivity and reversal differences occur because of difficulties with goal-directed learning dependent on top-down processes. However, attentional accounts suggest that this difficulty is relative to other attentional demands, and Newman and colleagues (Newman, 1998; Patterson & Newman, 1993) have suggested that psychopathic individuals are capable of processing peripheral or contextual information, but that this process is more effortful than it is in non-psychopathic individuals. This means that psychopathic individuals may or may not show altered learning and impulsivity depending on task parameters, and whether their attention is directed at the relevant cue (Newman et al., 1990).

In Budhani et al.'s (2006) experiment, the stimuli appeared as line drawings on a grey background, and the feedback as black text on the grey background. In comparison, in Gregory et al.'s (2015) experiment the stimuli were 2D, colourful images appearing on a black background, and the feedback was black text on a grey background, but also a green tick (correct) or a red cross (incorrect). Therefore, it could be argued that both the stimuli and the outcomes were more salient in Gregory et al.'s study, and so may have facilitated better goal-directed learning. But, if all the stimuli and outcomes were more

salient, then this bottom-up effect may have made it just as difficult for psychopathic individuals to focus on, and process the relevant cues. This explanation also conflicts with previous research, which suggests psychopathic individuals have difficulty with top-down attentional processes, and not bottom-up, salience driven ones (Hoppenbrouwers, Van der Stigchel, Slotboom, Dalmaijer, & Theeuwes, 2015).

Summary. So, whilst both outcome processing and attentional accounts may have their strengths, neither group seems to provide an adequate explanation of the data: outcome processing accounts do not explain why psychopathic individuals perform poorly on nonaffective tasks, or why they sometimes do not show learning and reversal differences (Hamilton et al., 2015); and, attentional accounts do not explain the situational specificity of psychopathy deficits and their suggested 'spotlight' approach is not considered broadly compatible with contemporary theories of attention (R. J. R. Blair & Mitchell, 2009; R. J. R. Blair, 2005). As such, the recent Impaired Integration Theory (Hamilton et al., 2015) was developed to reconcile and integrate these affective and attentional accounts of psychopathy, and in so doing made a bid to explain the diverse and complex literature on reversal learning.

1.4. Impaired Integration Theory

The Impaired Integration Theory (Hamilton et al., 2015) suggests that psychopathic individuals fail to rapidly integrate complex, multisensory information, which results in a perceptual bottleneck. This perceptual bottleneck leads to abnormal brain topography characterised by disrupted neural coordination. In a reciprocal fashion, these topographical and connectivity abnormalities hinder information integration. Therefore, the Impaired Integration Theory is dependent of four separable, but integrated processes:

1) *Perceptual Binding* – perceptual processing in psychopathy is narrow, meaning that attention is not extended to peripheral stimuli, and overall processing is superficial and limited.

2) *Learning* – shallow information processing can prevent the integration of past and present mental representations, meaning that prior learning cannot be evaluated and updated.

3) *Brain Topography* – the perceptual bottleneck shapes developmental connections and reduces 'use-induced' plasticity, leading to a topographical profile characterised by disrupted coordination.

4) *Emotion and Cognition* – the connectivity in emotion-related circuitry is underdeveloped, thereby reducing the propensity (not ability) to process peripheral information.

As such, the Impaired Integration Theory (Hamilton et al., 2015) suggests that learning in psychopathy is impaired due to superficial processing of stimuli that are the focus of attention, and virtually no processing of peripheral stimuli that fall outside the narrow attentional spotlight (e.g., Hoppenbrouwers et al., 2015). It is argued that the shallow information processing prevents the integration of past and present information, which impedes psychopathic individuals' ability to monitor and update S-R-O contingencies, leading to impulsivity and perseveration (e.g., Brazil, Maes, et al., 2013; Newman et al., 1987; Sadeh & Verona, 2008). The failure to process peripheral stimuli has been shown to be a relative difficulty, meaning that psychopathic individuals can process peripheral stimuli if this is possible within the confines of their narrow processing bottleneck, for instance in low perceptual load conditions, but not in high perceptual load conditions (e.g., Baskin-Sommers, Curtin, & Newman, 2013).

This relative difficulty processing peripheral information is also thought to account for psychopathic individuals' apparently poor processing of emotionally salient stimuli (Hamilton et al., 2015). Baskin-Sommers and colleagues (2013) showed the same pattern of impaired peripheral stimuli processing under high, but not low, load conditions when the peripheral stimuli were emotionally salient. This suggests that previous differences in processing affective information may, in fact, have reflected task conditions. As such, Hamilton et al. suggest that affective processing difficulties in high psychopathy individuals occur as a result of this bottleneck, and so affective information is not adequately processed if it is peripheral to the current goal focus, or if it requires complex processing. By suggesting that this difference is relative, the Impaired Integration Theory can again explain discrepancies between tasks reporting the presence or absence of affective processing difficulties in psychopathy (e.g., Baskin-Sommers et al., 2013; López et al., 2013; Munro et al., 2007a).

Therefore, it is argued that learning and reversal differences in psychopathy are relative, depending on the task demands and how these relate to the altered information processing capacity. For instance, in Budhani et al.'s (2006) study the effect of shallow information processing may have been particularly apparent because the stimuli and outcomes were distinguishable on only one dimension (i.e., shape), as compared to Gregory et al.'s (2015) study in which the stimuli and outcomes were distinguishable on two dimensions (i.e., colour and shape). Therefore, the Impaired Integration Theory (Hamilton et al., 2015) may be better able to account for the variability in learning and reversal differences that are seen in psychopathy.

Importantly though, the mechanism by which attention and learning interact is not specified within the Impaired Integration Theory (Hamilton et al., 2015) and so it is not clear why these differences occur. For instance, the impaired learning process is

described as resulting from shallow information processing, which prevents the integration of new information with previous learning. However, the information processing and learning process that is disrupted is not elaborated on. Likewise, the altered emotion and cognition process is said to reduce psychopaths' propensity, but not their ability, to process peripheral information. However, like the impaired learning process, the way in which these differences occur is not well outlined. As such, although these statements are useful in explaining the specificity of psychopathic abnormalities, such as impaired contingency monitoring under some conditions but not under others (e.g., Brazil, Maes, et al., 2013; Budhani et al., 2006; Gregory et al., 2015), they do not provide insight into the processes that cause these differences.

1.5. Learning Account of Psychopathy

An alternative approach that offers a mechanistic account of these differences involves interpreting the evidence of perseveration and reversal learning through a theory of learning, such as Mackintosh's (1975) associative account of attention. To date, there has been little attempt to harness theories of learning and attention as they have developed in the field of learning (although see Moul et al., 2012). In his 1975 Theory of Attention, Mackintosh formalised a theory of selective attention by establishing two principles of attention, gathered from an understanding of previous research. The principles were: 1) that we learn to attend to relevant stimuli and ignore irrelevant stimuli; and, 2) that rather than assuming a type of reciprocal rule, in which attention is simply what happens if one is not attending to something else (the so-called inverse rule; Thomas, 1970), Mackintosh suggested that the attention directed at a cue, and how easily it is learnt about, are determined by the extent to which a cue is uniquely predictive of its consequences. While the first principle seems quite uncontroversial the second principle

seems different in character from the bottleneck type models of attention proposed previously, whereby attention allocated to one stimulus restricts the attention allocated to another stimulus (for a review see Driver, 2001). Theories of selective attention have assumed that attention tunes in relevant stimuli at the expense of other stimuli, but Mackintosh allowed predictive or statistical information to determine attention.

Mackintosh (1975) proposed that these principles could be captured by a stimulus-specific learning-rate parameter: a measure of how easily a stimulus is learnt about, referred to as its associability, alpha (α). Previous theories of associative learning had focused on an error term (λ-V_A), used to represent the discrepancy between the expected outcome (V_A) and the actual outcome (λ; e.g., Rescorla & Wagner, 1972). Mackintosh's theory incorporated an error prediction term, but also included α and so suggested that associative change (ΔV) was governed by:

$$\Delta V_A = S\alpha_A (\lambda - V_A)$$

where A represents the cue being learnt about, V_A is the associative strength of stimulus A, S is a learning rate parameter, α_A is the associability of stimulus A and λ is the maximum conditioning possible to the outcome.

Unlike the Rescorla-Wagner model which used a summed error term (λ-$\sum V_A$; Rescorla & Wagner, 1972), Mackintosh's (1975) model used a separable error term for each cue meaning that it can only account for conditioned inhibition, in which an excitatory cue (e.g., X+) facilitates the development of inhibitory strength to another cue with which it is not reinforced in compound (e.g., AX-), with certain assumptions about the other parameters of the model. For example, inhibitory learning is predicted by assuming that lambda is determined by the nature of the bidirectionality of the outcomes. In animal studies the presence and absence of the outcome is represented by values of lambda of 1 and 0 (for the binary presence and absence). However, in many human

experiments the outcomes are continuous or part of a dichotomous experience (positive or negative). For instance, in a category learning experiment in which the outcome is an increase or decrease in a bidirectional outcome then lambda is logically defined symmetrically as +1 and -1 (Murphy, Schmeer, Vallée-Tourangeau, Mondragón, & Hilton, 2011), furthermore the effectiveness of these two outcomes might be different and represented by differences in the learning rate parameter (S). An alternative perspective was suggested by Le Pelley (2004) who incorporated a summed error term into his 'extended Mackintosh Model', given by:

$$\Delta V_A = S\alpha_A\,[\lambda-(\Sigma V-\Sigma \bar{V})]$$

where ΣV is the summed associative strength of all cues present, and $\Sigma\bar{V}$ is the summed associative strength of all presented stimuli for the US representation. This model is modified from the original to incorporate a summed error term, thereby retaining the strengths of the original Mackintosh model, whilst addressing the limitation of inhibitory association development. For the purpose of this thesis, this extended Mackintosh model will now be used in place of the original model.

The parameter α reflects the predictiveness of a cue, and so the correlation between the stimulus and an outcome. Importantly, this associability is relative: it depends in part on the associability of the other presented stimuli. This effect of other cues present is not, however, governed by the Inverse Hypothesis (Thomas, 1970), whereby a limited amount of attention is competed for by all the stimuli present, meaning that an increase in α to one stimulus is accompanied by a decrease in α to other stimuli. Instead, the impact of other stimuli present on a stimulus' α is related to its relative validity, or its unique predictiveness. For example, a stimulus that is moderately correlated with the outcome may have high associability if it is the best predictor of the outcome, but may have low associability if another cue present is a better predictor (Hall, Mackintosh,

Goodall, & Martello, 1977; Wagner, 1969). This is not because the presence of another cue 'uses up' some of the associability, but because the presence of another, more predictive cue, devalues the predictiveness of the original cue.

As such, α would increase for a stimulus that was well correlated with its consequences, and α would decrease if a stimulus were irrelevant, or poorly correlated with its outcome. Mackintosh (1975) suggested that this process of determining α is dependent on the following equations:

$$\Delta\alpha_A > 0 \text{ if } |\lambda - V_A| < |\lambda - V_X|$$

$$\Delta\alpha_A < 0 \text{ if } |\lambda - V_A| \geq |\lambda - V_X|$$

where V_X is the sum of the associative strength of all stimuli other than A present on that trial. As such, α_A will increase whenever the outcome of a trial is predicted better by A than by all other stimuli present, and α_A will decrease whenever the outcome of the trial is better predicted by another stimulus present. Therefore, Mackintosh used α to represent the idea that subjects learn to attend to relevant stimuli, and ignore irrelevant stimuli.

Evidence of this learnt associability has been seen in animals (Hall et al., 1977; Mackintosh & Holgate, 1968; Wagner, Logan, & Haberlandt, 1968). For instance, Hall et al. tested this effect using conditioned suppression training in rats. In Phase 1 all rats were conditioned to withhold their responding to a faint tone (A+). In Phase 2, half the rats continued to receive this pairing of the faint tone and the outcome (A+), whilst the other half had a salient light added (AB+). Although the groups were comparable at the end of Phase 1, at the end of Phase 2 the rats in the second group who received the compound training showed weaker suppression to the tone than the rats in the first group did. This indicated that when a more salient cue (the light) is paired with a stimulus, the previously predictive, but weakly salient cue (the tone) loses some of its relative validity,

and so its associability decreases. However, it is noted that this explanation makes the assumption that changes in α are able to influence behaviour, as well as learning, something not outlined in Mackintosh's original model.

This principle of stimulus competition, or the associative learning about stimulus-outcome relationships occurring relative to other learning, has been important in explaining other associative phenomenon. For instance, using the same procedure as that outlined in the Hall et al. (1977) experiment it is possible to produce blocking, whereby prior training with an element (A+) subsequently reinforced in compound with a novel cue (AB+) blocks the acquisition of associative strength about the novel cue (B; Kamin, 1969). Blocking was originally explained by suggesting that associative learning only occurs when an outcome is unexpected (Rescorla & Wagner, 1972). As A is established as a good predictor of the outcome in Phase 1, it is expected and unsurprising when it occurs in Phase 2, meaning that B acquires little associative strength. This is formally represented through the error prediction term, as the expected outcome ($\Sigma V_{AB} \approx 1$) will be very similar to the actual outcome ($\lambda = 1$).

However, Mackintosh (1975) argued instead that phenomenon like blocking can be explained in relation to the principle of relative predictiveness. In the case of blocking, a procedure dependent on the redundancy of the added element, Mackintosh argued that α_B would decrease as B signals no change from the expected reinforcement, and so with a low α, B would acquire very weak associative strength. But, as previously shown (Hall et al., 1977), were B more perceptually salient than A, it would have acquired associative strength. As such, Mackintosh was able to offer an explanation that did not depend on cues competing for associative strength or attention. Instead, by suggesting that it was changes in the cues' predictiveness and their relative perceptual salience that affects

learning, Mackintosh showed how this learning phenomenon could be explained in a model that determines cues' associative strength without the use of an inverse rule.

Mackintosh's (1975) principle of associability varying with a subject's experience (e.g., correlation between stimulus and reinforcement) has also been demonstrated in humans (e.g., Beesley & Le Pelley, 2011; Bonardi, Graham, Hall, & Mitchell, 2005; Le Pelley, Beesley, & Griffiths, 2014). For instance, Le Pelley et al. showed that humans can learn that stimuli are good predictors or poor predictors of their consequences, and that because of their higher associability the good predictors subsequently enter into new associations more easily than the poor predictors do. A cue (e.g., A) previously trained as predictive of outcome 1 (A1) was paired with a cue (e.g., B) that was poorly predictive of outcome 1 (B1) and the two were then trained in compound to predict a new outcome (2; AB2). The results showed that the previously predictive cue (A) acquired greater associative strength with outcome 2 than the previously non-predictive cue (B) did, despite both cues receiving the same number of pairings with outcome 2. Demonstrations such as this provide evidence that Mackintosh's theory governs attentional processes in learning.

So, it has been shown: how variations in a cue's associability can affect learning about the cue; and, how these processes are related to attention and the ability to ascertain the relative validity of a cue as a predictor of the outcome, a process dependent on integrating information across cues. It is possible that in psychopathy the information processing constraints outlined in the Impaired Integration Theory (Hamilton et al., 2015) affect the process of determining the associability of cues. This hypothesis would suggest two novel ideas about psychopathy, the first is that there is a general associability problem, which prevents a cue's associability from being updated, and the second is that previous evidence for an affective component to the disorder may simply reflect the

associability mechanism, and its insensitivity to adjustment on the basis of a cue's affective value. The principles of attentional accounts and outcome processing accounts that are incorporated in to the Impaired Integration Theory can be incorporated into an associative account (e.g., Mackintosh, 1975).

1.6. Explaining Reversal Inconsistencies

It might be possible to explain some of the inconsistencies in the psychopathic reversal learning literature using the associability account that I have just described. Though most of the reversal learning research has used either a go/nogo or an object discrimination procedure, variability in specific aspects of the methods may have influenced the associability of a cue, which is hypothesised to differently affect psychopathic and non-psychopathic individuals. For instance, the difference in results reported by Budhani et al. (2006) and by Gregory et al. (2015) may result from the effect that the varying cognitive loads had on participants' performance. Previous research has shown that the level of cognitive load has an effect on the type of learning that occurs (De Houwer & Beckers, 2003; Le Pelley, Beesley, & Griffiths, 2011; Le Pelley, Oakeshott, & McLaren, 2005), and as previously discussed, load affects the degree of perceptual processing that occurs in psychopathic individuals (Baskin-Sommers et al., 2013). As there were more cues and more trials in Budhani et al.'s experiment, the load of this task may be considered to be greater than that of Gregory et al.'s task. Therefore, it is possible that the additional trials and cues in Budhani et al.'s experiment adversely affected the psychopathic individuals, who already had a more limited processing capacity (Baskin-Sommers et al., 2013; Hamilton et al., 2015), making it particularly difficult for them to learn and monitor contingency changes.

Another methodological feature that may affect the associability of cues, and which applies exclusively to go/nogo tasks, is the use of an unequal number of go and nogo trials. Whilst many studies use a balanced number of cues that require responses and the withholding of responses (R. J. R. Blair et al., 2004; Brazil, Maes, et al., 2013; He et al., 2011; Kiehl et al., 2000), other studies do not apply such a balanced approach, and often give many more 'go' trials than 'nogo' trials (e.g., Lapierre et al., 1995; Newman et al., 1990). The imbalance in the frequency of trial types will affect the attention allocated to a stimulus and so its associability, meaning that more attention will be allocated to the stimuli that occur more frequently (Le Pelley et al., 2011; Mackintosh & Holgate, 1968), which is most commonly the excitatory trials that require a response. Therefore, in order to counteract this bias, more attention needs to be allocated to cues signalling the need to withhold a response, than to cues signalling the need to respond (Kiehl et al., 2000). This will make errors more likely in psychopathic individuals, who show altered attention (Newman & Baskin-Sommers, 2011; Wolf et al., 2012), and who are suggested here to show disrupted associability processes.

Reversal learning studies also differ in how response errors are interpreted, and how this relates to the contingencies governing learning. If a deterministic contingency is used then a cue is consistently paired with the outcome, and so any incorrect responses represent an error; whereas if a probabilistic contingency is used then there is a degree of inconsistency in the outcome that follow responses, and so the acceptable level of errors will depend on the probability of the contingency used. For instance, as previously discussed, Budhani et al. (2006) used both a deterministic contingency condition in which certain cues were always followed by a certain outcome, and a probabilistic contingency in which cues were followed by one outcome (e.g., reward) 80% of the time, and by a different outcome (e.g., punishment) 20% of the time. Therefore, in the

deterministic condition any errors violate the contingency, but in the probabilistic condition if there is a 20% error rate in responses to an inhibitory cue, so a 20% rate of errors of commission, this is permissible as it is consistent with the reinforcement received by participants.

However, Budhani and colleagues (2006) used a deterministic interpretation of response errors in both the deterministic and probabilistic conditions, thereby overestimating the number of errors made by individuals in the probabilistic condition. This study reported that psychopathic individuals showed reversal problems, and were worse at learning information that involves uncertainty, or a more subtle change in reinforcement (R. J. R. Blair, 2005). In fact, psychopathic individuals' higher error rate in the probabilistic condition as compared to the deterministic condition suggests that they were more accurate and showed a greater sensitivity to the subtleties of the contingency than their non-psychopathic controls did. Indeed, psychopathic samples have been shown to be capable of distinguishing 100-0, 80-20 and 50-50 contingencies (von Borries et al., 2010). These examples of cognitive load, go/nogo trial imbalances and response error interpretation show how differences in methodological features can affect the associability of a cue, or even more directly affect the results of studies designed to assess learning and contingency monitoring in psychopathy.

1.7. Assessing Inhibition

As discussed, reversal learning requires participants to monitor the contingency between stimuli, responses and outcomes, and how these contingencies change across a series of trials, or phases. As well as happening on this relatively large scale, such contingency monitoring, and to some extent reversal, occurs on a much smaller scale from one trial to the next, particularly when learning about inhibitory cues that require

the integration of information across stimuli. So, how a cue is trained will affect the associations it forms, and so the type of contingency monitoring that is required. This highlights: the distinction between learnt, or associative inhibition, and training that might result in the reduced or weakened likelihood of responding but without the contribution of inhibition (i.e., through habituation or learned inattention); and, the different effects these processes have on contingency monitoring and a cue's associability (McLaren & Verbruggen, 2015). For example, V trained by itself with the absence of the outcome (V-) might be unlikely to elicit a response, but the lack of responding is not evidence of inhibitory control or an inhibitory association between V and the outcome (McLaren & Verbruggen, 2015). Stronger inhibition is trained when an inhibitor (e.g., A) is trained in the presence of a signal for an outcome (e.g., X+) but with the outcome omitted (e.g., X+, AX-; Mackintosh, 1975; McLaren & Verbruggen, 2015; Rescorla & Wagner, 1972; Rescorla, 1969).

This stronger inhibitory training comes from monitoring the conflicting X+ and AX- trials (McLaren & Verbruggen, 2015; Rescorla, 1969). Under these conditions, there will be an association formed between X and the outcome, and when the outcome does not occur following the presentation of the compound AX, there will be a discrepancy between the actual and the expected outcome. This disparity is represented by a prediction error, such as that previously outlined from the Rescorla-Wagner model (1972; $\lambda - \Sigma V_{AX}$) where λ represents the actual outcome, and ΣV represents the sum of the expected outcome given all the cues present. Therefore, if there is an expectation that the outcome will occur ($\Sigma V_{AX} = 1$), but the outcome does not occur ($\lambda = 0$), the prediction error will be negative (-1), meaning that A acquires inhibitory associative strength. As A is predictive of the absence of the outcome, its associability will also likely increase (Mackintosh & Holgate, 1968; Mackintosh, 1975). In contrast a cue that

is unlikely to elicit a response, but is not an inhibitor (e.g., V-) will likely have an associative strength near zero and a reduced associability, as it is not predictive of either the presence or the absence of the outcome.

Which one of these two procedures is followed will affect the strength of the association, the cue's associability, and how the cue is subsequently learnt about. The retardation, or disruption of reversal learning will be significantly greater for an associative inhibitor than for a cue associated with a reduced tendency to respond without the involvement of inhibition, or for a cue that has not been learnt about (Rescorla, 1969, 1971a). The difference between inhibitors, neutral cues or cues that have not been learnt about may not be apparent when the cues are directly tested, as all three will result in the absence or withholding of a response. As such, to distinguish these different types of cue Rescorla (1969) argued that associative inhibition must be tested against an excitatory background, for example by measuring how an inhibitory cue is re-trained to become an excitor.

This process of re-training, known as a retardation test, is akin to the reversal training frequently used in the psychopathy literature. Differences between psychopathic and non-psychopathic individuals are often reported only in this reversal phase (e.g., Budhani et al., 2006; Newman et al., 1987), which emphasises the need to test cues against an excitatory background as the differences are not always evident in the acquisition phase. Typically these differences manifest as psychopathic participants showing weaker reversal, and so making more errors than their non-psychopathic controls in the reversal phase, results seen as evidence of poor inhibition (e.g., Budhani et al., 2006). However, according to associative predictions such as those outlined by Mackintosh (1975) or Rescorla (1969): if inhibition has been acquired, then participants' ability to shift this association to become more excitatory will be impaired, or retarded,

as compared to a cue that did not acquire the initial inhibitory association. Therefore, reduced reversal could be seen to indicate stronger initial inhibitory learning. Mackintosh's (1975) theory predicts a certain degree of reversal learning as the cue's associability will be high following the acquisition phase, and so this will initially facilitate the reversal learning. However, as the trials progress the inhibitor's α will decrease, and so learning will diminish, thereby limiting the extent of total reversal that is seen.

It has been suggested that the degree to which this retardation occurs also depends on the number of trials participants are exposed to, and whether the reinforcement is probabilistic, meaning that there is a degree of variability about the reinforcement a cue receives (Mackintosh & Holgate, 1968). Mackintosh and Holgate showed that as both the number of consistently reinforced trials, and the consistency of the reinforcement increases, so does the strength of attention allocated to a cue. Therefore, if psychopathic individuals' associability is disrupted, then they may be particularly affected by these factors due to an exaggerated problem in allocating sufficient attention to the necessary cues. Support for this suggestion is seen in psychopathic individuals having a lower level of accuracy (i.e., more response errors) after one block of training, but their accuracy being comparable to controls by the fourth block of training (von Borries et al., 2010).

In addition to the retardation test, Rescorla (1969) proposed that inhibitors can be tested using a summation test in which the propensity of an inhibitor (e.g., A) to suppress responding to an excitatory cue (e.g., Y+) when they are presented in compound (e.g., AY+) is compared to responding to an excitatory cue (e.g., D+) when it is presented in compound with a novel cue (e.g., DS+). If the test cue acquired inhibitory strength, it is predicted to reduce responding to the excitatory cue that it is presented with. He et al. (2011) used such a summation test and found that psychopathic offenders, though

appearing to show inhibition in the acquisition phase, did not show summation, instead making equivalent responses to the test compound (e.g., AY+) and the comparison compound (e.g., DS+). Non-psychopathic offenders did show summation. Again this shows the importance of testing inhibitory associations against an excitatory background.

An alternative approach to assessing contingency monitoring and inhibition is to indirectly assess the impact of inhibitory learning on excitatory processes. For example, super-learning is a particularly important prediction of models of active inhibition (Rescorla, 1971b), and has previously been reported in animals (B. A. Williams & McDevitt, 2002; D. A. Williams, Mehta, & Dumont, 2004) and in humans (Lovibond, Siddle, & Bond, 1988). Super-learning is the facilitation of excitatory learning about a novel cue (e.g., Y) that occurs when that novel cue is reinforced in compound with an inhibitor (e.g., AY+; Rescorla, 1971b). Models of learning based on error prediction suggest that this enhanced learning and excitatory association about Y is a reflection of the error term: $(\lambda - \Sigma V)$. As outlined above λ represents the actual outcome, and ΣV represents the sum of the expected outcome. According to this, if the expectation is negative (as follows from inhibitory learning; $\Sigma V_{AY} = -1$), and the actual outcome is positive ($\lambda = 1$), the error prediction term will be large, as compared to a novel cue (e.g., DS+), with no expectations ($\Sigma V_{DS} = 0$) that is paired with the positive outcome ($\lambda = 1$). Therefore, if the participants have acquired an inhibitory association, super-learning will be possible, but, if they have not, then super-learning will not occur. Like the retardation and summation tests, the use of super-learning provides a way of testing inhibition against an excitatory background.

1.8. Novel Method for Assessing Learning in Psychopathy

This thesis presents a series of experiments that were conducted to assess learning in relation to psychopathic traits, and to test the role of associability in this learning. Chapters 2-4 develop a novel go/nogo paradigm, and test how this can be used to measure contingency monitoring through the acquisition and shifting of inhibitory associations, and the conditions under which this learning varies. I sought to test: whether associative inhibition (AX-, X+) could be acquired; how these associations were shifted to changes in the stimulus-outcome contingency; whether this inhibition could be used to induce enhanced (super) learning about a second cue; whether these effects were symmetrical about excitatory and inhibitory cues; and, if psychopathic traits predicted differences in these effects.

In addition, separate assessments of participants' ability to acquire stimulus-response (S-R) and stimulus-outcome (S-O) associations were also made by measuring both the accuracy of participants' responses (S-R), and by requiring them to make ratings of the inhibitory or excitatory strength they perceived a stimulus to have (S-O). This type of rating measure has previously been shown to be effective in assessing the strength of associative learning (Byrom & Murphy, 2013; He et al., 2011). By measuring both ratings and responses I was able to contrast participants' ability to form associations between stimuli and the outcome, so-called goal-directed behaviour (ratings), with their ability to form stimulus-response associations, so-called habit-like behaviour (responses; De Wit & Dickinson, 2009). These two distinct processes are confounded by studies that only measure participants' response accuracy, and infer from this whether or not participants were able to form the necessary associations (R. J. R. Blair et al., 2004; Budhani et al., 2006; Munro et al., 2007b). Yet, the distinction in these goal-directed and habit-like abilities is central to a number of accounts of psychopathy (e.g., Gorenstein & Newman, 1980; Moul et al., 2012).

This distinction in abilities is developed in Chapter 5, where an alternative previously established method for assessing perceptions of instrumental control was used (Byrom, Msetfi, & Murphy, 2015). This instrumental task provided an independent test of the predictions generated following the results of Experiments 1-6, and an alternative means of assessing participants' contingency monitoring, by contrasting a relatively stable, predictive contingency with one that was ambiguous and constantly reversing. By focusing on a direct test of contingency learning, it was also possible to directly test participants' processing and learning about contextual information.

I tested whether psychopathic traits predicted a difference in performance on these tasks. Though psychopathy is a good predictor of illegal behaviour (Salekin, Rogers, & Sewell, 1996), since its early clinical conceptualisation it has been emphasised that criminal behaviour is neither a sufficient nor necessary feature of psychopathy (Cleckley, 1982). This means that psychopathy can be studied in non-criminal samples, such as University students, or members of the general population (Z. Lee & Salekin, 2010), as well as forensic samples. For instance, a study comparing high and low psychopathy community samples on the Iowa Gambling Task found the same pattern of results that is seen in high and low psychopathy forensic samples (Mahmut, Homewood, & Stevenson, 2008). Following results such as this, it has been suggested that forensic and community psychopathy samples both possess the characteristic interpersonal-affective and antisocial features of psychopathy, but that these manifest to a more extreme degree in forensic samples (Benning, Patrick, Blonigen, Hicks, & Iacono, 2005; Z. Lee & Salekin, 2010; Mahmut et al., 2008). As such, mixed community samples of students and non-students were considered an appropriate means to test whether psychopathic traits predict differences in learning and inhibition.

CHAPTER 2

Assessing Learning in Psychopathy

(Experiment 1)

2.1. Introduction

Chapter 1 provided a review of the literature related to reversal learning differences in psychopathy, and highlighted the degree of variability in these reported differences. Attentional (e.g., Gorenstein & Newman, 1980) and outcome processing (e.g., R. J. R. Blair, 2005) accounts, the primary explanations for these results, were discussed as was the Impaired Integration Theory (Hamilton et al., 2015): a model designed to incorporate both attentional and outcome processing principles as well as a neurobiological perspective. I argued that the processes outlined in these models are governed by associative learning (Mackintosh, 1975; Pearce & Hall, 1980; Rescorla & Wagner, 1972), and that differences in how associations are learnt may, therefore, underlie differences in psychopathy. As such, Experiment 1 reports the development of a novel variant of a go/nogo task designed to test if associative procedures can be embedded in a go/nogo task, and whether psychopathic traits predict differences in associative learning, with specific focus on the role of associative inhibition.

Behavioural and cognitive processes were separately assessed through the accuracy of timed-responses to stimuli (behavioural learning) and untimed explicit causal judgements of stimuli (explicit learning). As it has been suggested that poor inhibition underlies impulsivity in psychopathy (He et al., 2011; Kiehl et al., 2000;

Newman et al., 1985), participants' ability to learn about inhibitory cues, which signal the need to withhold responding, was tested. This was done both directly and indirectly by measuring: learning about the inhibitor itself; the ability to flexibly update this learning following stimulus-outcome contingency changes (shift learning); and, whether the inhibitor had an effect on learning about other cues, as measured through the super-learning effect described in Chapter 1 (1.7.; Rescorla, 1971b; B. A. Williams & McDevitt, 2002).

As such, participants' learning was evaluated through three calculated variables: *shift errors*, the difference in response accuracy to a cue following contingency changes between the acquisition phase and the shift phase; *shift ratings*, the difference in ratings to a cue following contingency changes between the acquisition phase and the shift phase; and, *super-learning*, the difference in ratings between the super-learnt cue and a comparison cue in the shift phase. These measures were incorporated into a design that balanced the number of go and nogo trials and so balanced exposure to excitatory and inhibitory stimuli. This novel method was applied to a community sample of students and non-students to assess the relationship between psychopathic traits and inhibitory learning.

Two of the three subscales of the Psychopathic Personality Inventory Revised (PPI-R; Lilienfeld & Widows, 2005) were focused on: Fearless Dominance (FD) and Self-Centred Impulsivity (SCI). Coldheartedness (C), the third sub-scale, which measures 'meanness', is poorly correlated with FD and SCI and was not expected to predict performance on this type of non-social task (S. O. Lilienfeld & Widows, 2005). FD reflects the primary psychopathy sub-type, centred on social and physical boldness, and a relative immunity to anxiety or guilt (Benning et al., 2003; Hare, 1991; S. O. Lilienfeld et al., 2012). Experimentally it has been shown to predict altered control and a

reduced sensitivity to (negative) feedback (Benning et al., 2003; Patrick et al., 2009). FD also represents the agentic qualities of psychopathy, such as social potency, and has been suggested to represent both the positive and the negative aspects of psychopathy (Patrick et al., 2009). For this reason, Patrick and colleagues have argued that the construct of FD is likely to be particularly relevant when studying psychopathy in community (i.e., not forensic) samples (Patrick et al., 2009). Given its relation to altered feedback sensitivity, FD was predicted to be associated with disrupted learning and responding.

Previous research has suggested that response patterns can influence peoples' perceptions of the relationship between stimuli (Matute, 1996). Specifically, individuals who respond more often tend to over-estimate the relationship between their actions and the occurrence of an outcome, as compared to individuals experiencing the same response-outcome contingency but who respond less often. Given this effect of behaviour on perceptions of control, I used mediation analysis to test the direct relationship between FD and super-learning, and whether this relationship was mediated by shift errors and/or shift ratings. This was done because shift errors may have influenced the ratings of the shift cue (Matute, 1996), and because the shift cue was used to facilitate super-learning. These potential relationships can be seen in Figure 1.

SCI represents psychological and behavioural maladjustment (S. O. Lilienfeld & Widows, 2005; Patrick et al., 2009) and is associated with increased impulsivity. SCI is suggested to represent the secondary psychopathy sub-type, which like the primary sub-type is characterised by anti-social behaviour, but importantly also reflects poor impulse control, as well as a preserved capacity to feel guilt, empathy and anxiety (S. O. Lilienfeld et al., 2012). As such, SCI was predicted to relate to increased behavioural effects, demonstrated through greater shift errors (errors in the shift phase relative to errors in the acquisition phase). As response behaviour can influence learning about

contingencies (Matute, 1996), it follows that SCI may indirectly affect super-learning via its affect on shift errors, which in turn influences shift ratings, on which super-learning is based. These potential relationships are also shown in Figure 1.

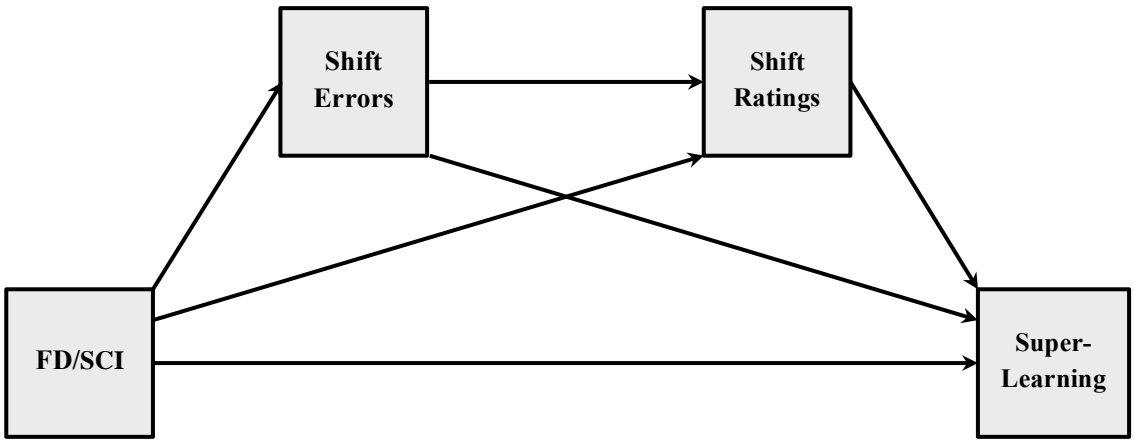


Figure 1. The mediation model of the hypothesised direct effect between either FD or SCI and super-learning, and the indirect effects through the potential mediators of shift errors and/or shift ratings.

2.2. Method

2.2.1. Participants

Seventy-eight (20 male, 58 female; age 18-24 years) non-colour-blind participants were recruited and given course credit or payment. The number of participants was determined using a power analysis of pilot data relating to the shift in learning about the inhibitory test cue (A). On the basis of the pilot data ratings in the acquisition and shift phases ($\mu_1 = 3.04$; $\mu_2 = 3.75$; $\sigma = 1.58$) a power analysis indicated that 78 participants were needed to have 80% power for detecting an effect when employing the traditional .05 criterion of statistical significance. Ethical approval for the study was granted by the

University of Oxford Research Ethics Committee, reference number MSD-IDREC-C2-2013-008.

2.2.2. Design

Shift learning (shift errors and shift ratings) and super-learning were both assessed using 2 x 2 mixed model designs. For the shift learning, the within-subject variable was the phase (acquisition vs. shift), and for the super-learning the within-subject variable was the cue type (super-learnt vs. comparison). In both cases two stimuli counterbalance conditions and sex were included as between-subjects variables. Psychopathic traits were separately included as independent variables in the analysis.

2.2.3. Materials

Psychopathic Personality Inventory (Revised; PPI-R; Lilienfeld & Widows, 2005). This self-report questionnaire assesses psychopathic traits through 154 items, 64 of which are reverse scored. Participants rate the items as: false, mostly false, mostly true or true. The PPI-R divides in to three factors: Self-Centred Impulsivity, consisting of the Machiavellian Egocentricity, Rebellious Non-Conformity, Blame Externalisation, and Carefree Non-Planfulness subscales; Fearless Dominance, comprising the Social Influence, Fearlessness and Stress Immunity subscales; and, Coldheartedness. *T*-transformations were carried out on the Total PPI-R scores and the factors' scores on the basis of sample type (community vs. forensic), sex and age group of the participant. In the present sample, the internal consistency of the PPI-R was good ($\alpha = .92$). Participants' means and standard deviations are shown in Table 1.

Table 1

Experiment 1 Participant Psychopathy *T*-Scores.

	M	*SD*
Total PPI-R	48.15	9.84
Self-Centred Impulsivity	50.59	8.61
Fearless Dominance	46.06	9.48
Coldheartedness	49.87	10.43

Note: N = 78. The mean *t*-score is 50.

Shift Go/NoGo Task. Participants were required to learn when to respond and when to withhold their responses to ten different cues that had positive (excitatory: X+, AY+, BV+, DS+) or negative (inhibitory: E-, V-, AX-, CZ-) contingencies, as shown in Table 2. Participants were required to press a key to indicate the prediction that the stimulus shown would increase sales, and to withhold their responding to indicate a prediction that the stimulus shown would decrease sales. This type of bidirectional outcome (as opposed to the unidirectional presence or absence of an outcome) has previously been shown effective when training excitatory and inhibitory associations (Baetu & Baker, 2010; Melchers, Wolff, & Lachnit, 2006). For instance, Baetu and Baker show the same pattern of conditioned inhibition and summation when using a bidirectional or a unidirectional outcome. There were 120 trials in the acquisition phase (30 per stimulus) and 60 trials in the shift phase (10 per stimulus). Participants' decision to respond (or not) to the stimuli determined response errors (commission or omission), whilst their explicit learning was assessed through cue-ratings they made at the end of each phase. This type of category learning task has previously been shown to produce good learning (Byrom & Murphy, 2013; He et al., 2011). As shown in Table 2, the

presentation of excitatory and inhibitory cues was balanced within each phase to ensure that any shift effects were not due to participants only having received inhibitory cues in the acquisition phase and excitatory cues in the shift phase, or vice versa.

Included in the design was a test cue, A, that had its associations shifted between the two phases. A was trained as an inhibitor in the acquisition phase, meaning it was predictive of the absence of the outcome (AX-, X+; Rescorla, 1969). As discussed in Chapter 1 (1.7.), this type of associative inhibitory training is different from simply pairing a cue with the absence of the outcome, and consequently results in much stronger inhibitory learning (McLaren & Verbruggen, 2015).

Table 2

Experiment 1 Design

Acquisition Phase	**Shift Phase**
AX-, X+	AY+, AX-, X+
BV+, V-	CZ-
	DS+
	E-

Note: Overall outcome direction is shown as excitatory (+) or inhibitory (-). A is the shift test cue; Y is the super-learnt cue (to be compared to S). All other cues are included to counterbalance the training of A and Y.

The association between A and the outcome was then modified in the shift phase (AX-, AY+). The total response errors to A in the acquisition (Aq) and shift (S) phases

were compared to determine whether shift learning occurred ($AError_{Aq} = AError_S$) or not ($AError_{Aq} < AError_S$); and, likewise the ratings of A in the acquisition and shift phases were compared to determine whether shift learning occurred ($ARating_{Aq} < ARating_S$) or not ($ARating_{Aq} = ARating_S$).

Test cue A was predicted to facilitate super-learning (B. A. Williams & McDevitt, 2002) about cue Y with which it was trained in compound in the shift phase (AY+). Having previously been trained as an inhibitor, A was predicted to facilitate enhanced excitatory learning about Y when reinforced in compound due to the large discrepancy between the predicted ($\Sigma V_{AY} = -1$) and the actual ($\lambda = +1$) outcome, as captured through the error prediction term outlined in Chapter 1 (1.5.; $\lambda - \Sigma V_{AY} = +2$; Rescorla & Wagner, 1972). Y was compared with a novel excitatory cue also trained in compound (S) to determine if super-learning had occurred (Y > S) or not (Y = S). This provided a second-order test against an excitatory background of the inhibitory strength acquired by A in the acquisition phase (Rescorla, 1969).

The stimuli were consistent with the selected cover story of monitoring ice cream sales, meaning that each cue was represented by a flavour (both the colour of the 'scoop' and the name were shown). The trial sequence is shown in Figure 2. Each trial began with a three-second countdown. Following this, there was the presentation of a stimulus, which was either an element (one ice cream flavour) or a compound (two ice cream flavours). The maximum stimulus presentation time was 1000ms, though the cues were removed and the trial progressed if participants responded (by pressing the 'b' key).

After a response was made, or the maximum time reached, participants were shown feedback for 1250ms that was contingent on their response accuracy. The feedback was either: "RIGHT, sales did INCREASE today!"; "RIGHT, sales did DECREASE today!"; "WRONG, sales actually DECREASED today!"; or, "WRONG,

sales actually INCREASED today!". The feedback was probabilistic, occurring in one direction on 80% of trials for a given cue. For example, in the shift phase the compound CZ received eight inhibitory trials, and two excitatory trials. At the end of each phase participants were shown the cues one at a time and asked to rate them on a scale from 1 (inhibitory; reduced ice cream sales) to 10 (excitatory; increased ice cream sales).

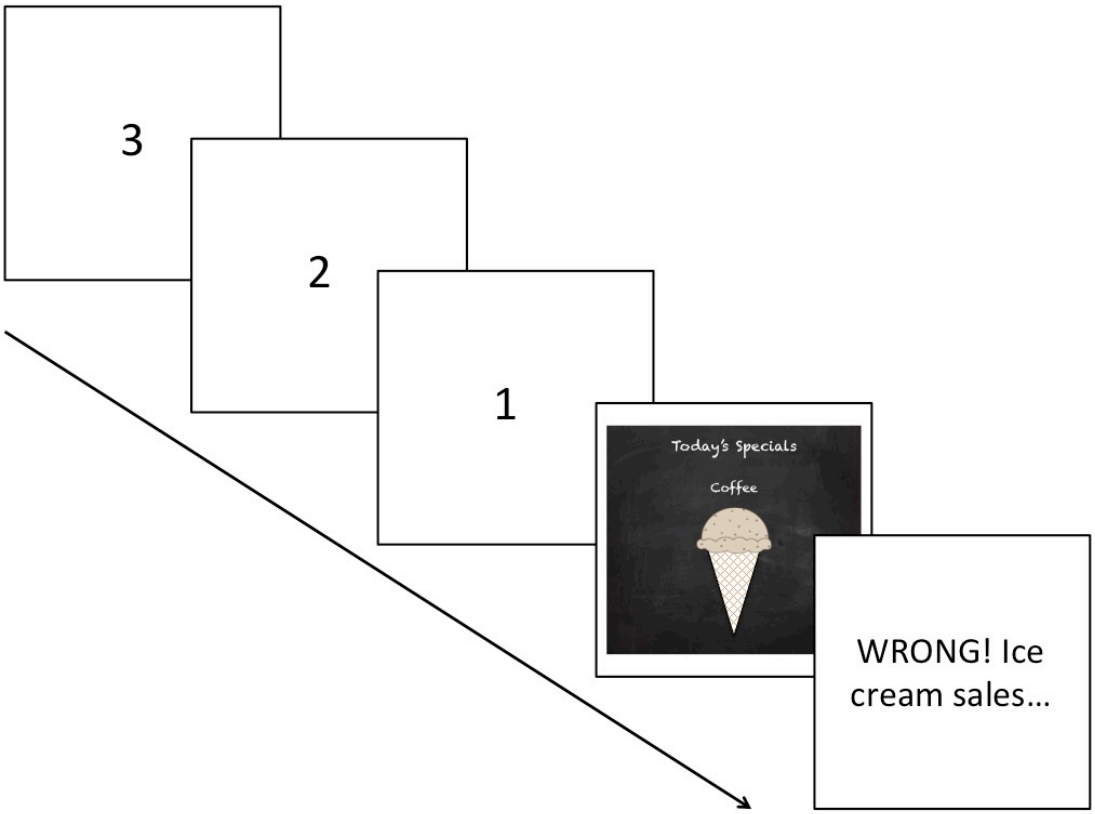


Figure 2. Experiment 1 example trial sequence.

The stimuli were counterbalanced across two different conditions to which participants were randomly allocated. Two counterbalance conditions were used so that this could be included as a variable in the statistical analysis to ensure that any of the observed effects were not due to the flavour of ice cream associated with a cue, for example A always being represented by caramel and B by chocolate. Within these two

counterbalance conditions the following cues were counterbalanced: A – X, B – V, Y – S. So, for instance, two example conditions may appear as follows:

Flavour Condition 1	Cue	Flavour Condition 2
Chocolate	A	Caramel
Caramel	X	Chocolate
Strawberry	B	Coffee
Coffee	V	Strawberry
Vanilla	Y	Pistachio
Pistachio	S	Vanilla

It is noted that A and B are not counterbalanced, and as such any differences between them must be interpreted in the light of this potential confound.

2.2.4. Procedure

Having given their consent, participants first completed the go/nogo task, then the PPI-R (S. O. Lilienfeld & Widows, 2005) before being debriefed.

2.2.5. Statistical Analysis

Initial acquisition of associations was tested in the acquisition phase by comparing the ratings of the inhibitory cue (A) with an excitatory cue (B). This was done using a mixed model analysis of variance, with cue type as the within-subjects factor, and sex and stimulus counterbalance condition as the between-subjects factors. Sex was included in the analysis given previous evidence of differences in psychopathic traits between males and females and how these traits relate to external variables (Cale & Lilienfeld, 2002; Lilienfeld & Hess, 2001). For instance, Lee & Salekin (2010) administered the PPI-R Short Form to a large sample of students (n = 1229) and found that factor loadings differed significantly between male and female participants. Studies such as this have led to discussion around sex differences in psychopathy (Blonigen & Carlson,

2003; Cale & Lilienfeld, 2002) and to the inclusion of sex in analysis of psychopathic traits' relationship to external variables, such as fear-conditioning (e.g., Gao, Raine, Venables, Dawson, & Mednick, 2010).

Following this, I assessed the dependent measures: shift errors; shift ratings; and, super-learning ratings. Shift errors were calculated as a proportion of the total possible trials on which an error to the test cue (i.e., A) could be made, and this difference score was calculated by subtracting the proportion of errors made in the shift phase, from those made in the acquisition phase. This facilitated comparison between the acquisition phase, in which there were 30 trials of each cue, and the shift phase, in which there were 10 trials of each cue. Comparison of the raw scores would have overestimated the number of errors made during the acquisition phase. Shift ratings were also calculated as a difference score by subtracting the rating of the inhibitory cue (A) in the acquisition phase, from its rating in the shift phase, meaning that a positive score represents a greater shift in ratings. Super-learning rating was calculated by subtracting the rating of the comparison cue from the rating of the super-learnt cue in the shift phase, meaning that a positive score represents greater super-learning.

Mixed model analyses of variance (ANOVA) were used to assess the shift errors and shift ratings in relation to the within-subjects factor of phase (acquisition vs. shift). Sex and flavour counterbalancing conditions were included as between-subjects variables. A mixed model ANOVA was also used to assess super-learning ratings with respect to the within-subjects variable of cue type (super-learnt vs. comparison). Sex and flavour counterbalancing conditions were included as between-subjects variables. *Post hoc t*-tests were only conducted following significant results of ANOVA and were corrected using Bonferroni.

Zero-order correlations were calculated between each of the dependent variables, and the overall psychopathy score (total PPI-R score), in addition to the PPI-R factors of SCI, FD and C. Multiple mediation analyses with mediators operating in serial were conducted to test whether either FD or SCI predicted super-learning, and whether these relationships were mediated by shift errors or shift ratings. As outlined in the introduction (2.1.), shift errors were included as a mediator variable because responses can alter learning about contingencies (Matute, 1996), which is this case might affect the shift ratings. Shift ratings were included as a mediator variable because the super-learning was based on the inhibitory value of the shift cue, therefore it was predicted that the extent of shift learning would inversely predict super-learning. These relationships were tested using a non-parametric bootstrapping procedure (PROCESS; Preacher & Hayes, 2008) which estimates the indirect effects of multiple mediator variables, and in this case was based on 1000 bootstrap resamples with 95% bias corrected confidence intervals. Sex and flavour counterbalance conditions were controlled for in the models. An alpha level of .05 was used throughout.

2.3. Results

Participants made significantly more errors in the shift phase than they did in the acquisition phase. Participants' ratings of the test cue (A) showed significant shift learning, and as shown in Figure 4 female participants, but not male participants, showed super-learning. Neither FD nor SCI predicted any of these effects.

2.3.1. Overall Learning Effects

Initial Cue Learning (Acquisition Phase). As shown in Figure 3 participants

rated the inhibitory cue (A; $M = 3.20$, $SD = 1.67$) as significantly more negative than the excitatory cue (B; $M = 6.82$, $SD = 1.48$), $F(1,74) = 17.51$, $p < .001$, $\eta_p^2 = .64$, 95% CI (.50, .72). No interactions were significant, all Fs < .31, all ps > .58.

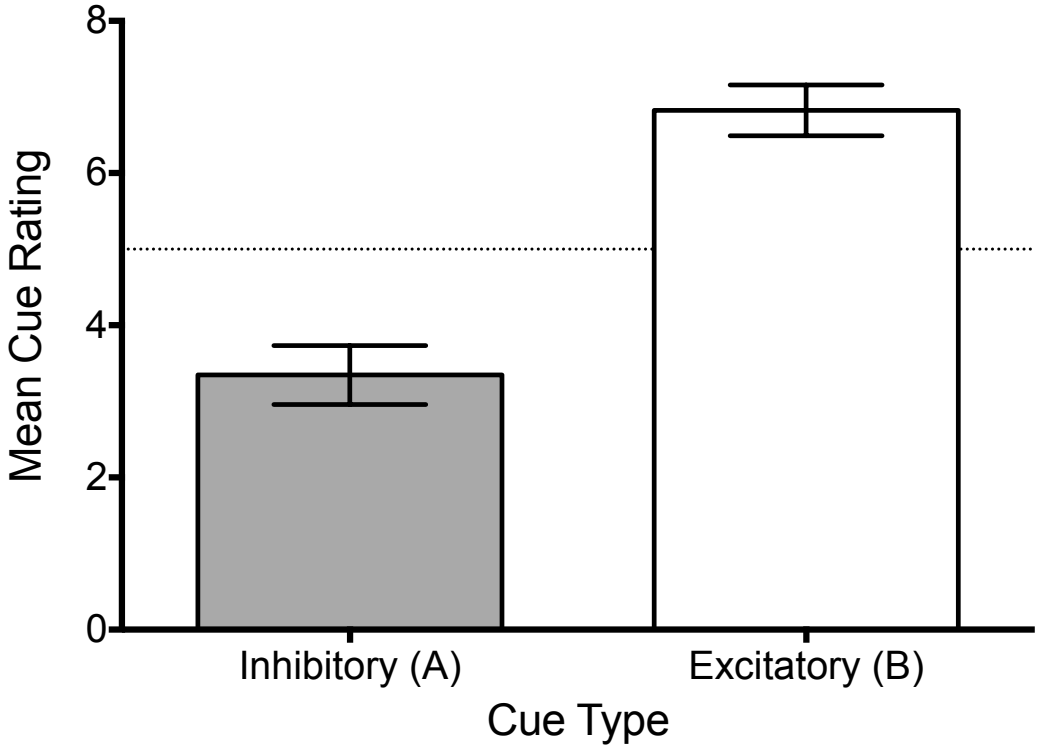


Figure 3. Participants rated the inhibitory (A) and the excitatory (B) cues as significantly different from each other in the acquisition phase. The rating scale ran from 0-10, where 0 represented a strong inhibitor, and 10 a strong excitor. Error bars represent 95% confidence intervals.

Shift Errors. Participants made a significantly greater proportion of errors in the shift phase ($M = .38$, $SD = .28$) than they did in the acquisition phase ($M = .26$, $SD = .21$), $F(1,74) = 130.93$, $p < .001$, $\eta_p^2 = .19$, 95% CI (.05, .34). No interactions were significant, all Fs < .98, all ps > .32.

Shift Ratings. Participants rated the target cue (A) significantly more positive in the shift phase (M = 3.58, SD = 1.58) than in the acquisition phase (M = 3.20, SD = 1.67), $F(1,74) = 6.70$, $p = .012$, $\eta_p^2 = .08$, 95% CI (.01, .22). No interactions were significant, all Fs < 1.92, all ps > .72.

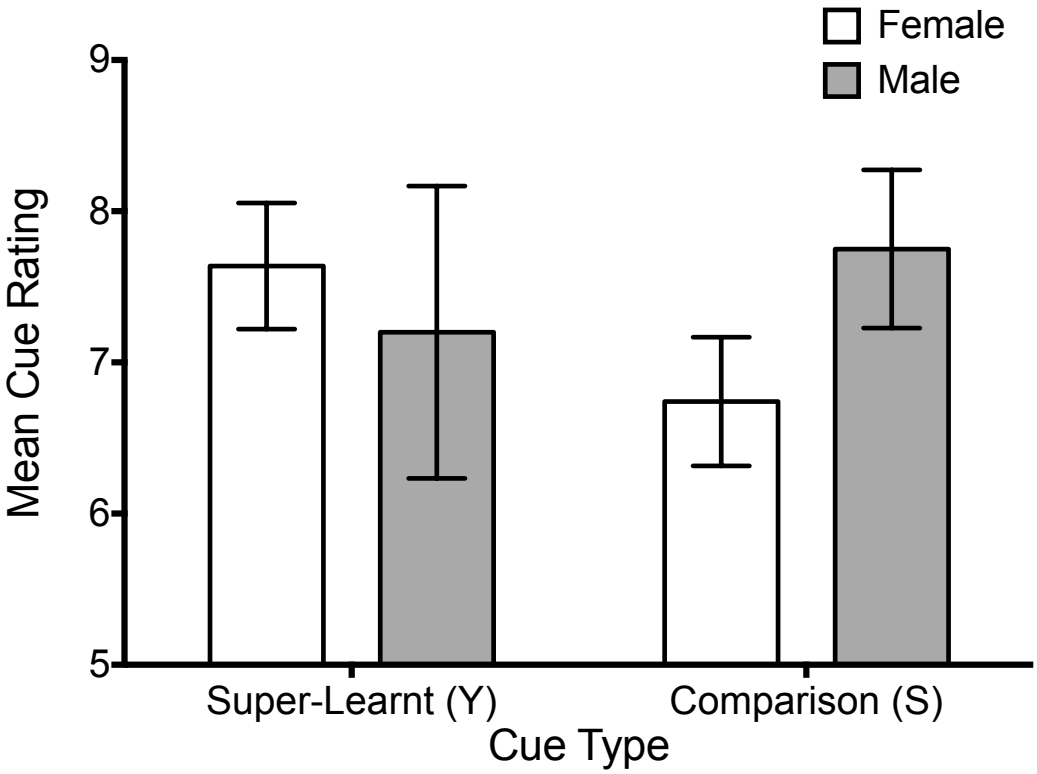


Figure 4. Female participants rated the super-learnt cue (Y) significantly more positive than the comparison cue (S), however, male participants did not. Error bars represent 95% confidence intervals.

Super-Learning. As shown in Figure 4, there was a significant interaction between cue type and sex, $F(1,74) = 7.99$, $p = .006$, $\eta_p^2 = .10$, 95% CI (.01, .23). *Post hoc* paired samples *t*-tests showed that whilst the male participants did not rate the super-learnt cue (M = 7.20, SD = 2.07) and the comparison cue (M = 7.75, SD = 1.12)

significantly differently, $t(20) = -.98$, $p = .342$, $\eta_p^2 = .05$, 95% CI (.00, .28), the female participants did (super-learnt cue: $M = 7.64$, $SD = 1.59$; comparison cue: $M = 6.74$, $SD = 1.62$), $t(57) = 3.50$, $p < .001$, $\eta_p^2 = .18$, 95% CI (.03, .34). No other main effects or interactions were significant, all Fs < 4.10, all ps > .47.

2.3.2. Relationship Between Psychopathy and Dependent Variables

The zero-order correlations between overall psychopathy score, the factors of SCI, FD and C, and the dependent variables are shown in Table 3. Total PPI-R score was negatively related to shift errors ($r = -.23$, $p = .037$). All other dependent variables were not significantly correlated with psychopathy traits.

Table 3

Experiment 1 Zero-Order Correlations

	Total PPI-R	**SCI**	**FD**	**C**
Total PPI-R	-			
SCI	**.804*** **	-		
FD	**.793*** **	**.324** **	-	
C	**.501*** **	.215	**.329** **	-
A Shift Errors	**-.234***	-.212	-.151	-.117
A Shift Ratings	-.012	-.054	-.028	-.181
SL	.071	.121	-.039	.037

Note: Total PPI-R, SCI, FD and C represent the *t*-scores for each variable. SL = Super-Learning Ratings.

*** $p < .001$; ** $p < .01$; * $p < .05$.

The mediation analysis showed that positive super-learning was not directly or indirectly predicted by FD [all βs < .01, all 95% CIs (< -.01, > .01)], or by SCI [all βs < .04, all 95% CIs (< -.01, > .01)].

2.4. Discussion

These results show that psychopathic traits in a community sample are not predictive of response impairments, or the inability to learn about inhibitory stimuli as measured either directly (shift ratings) or indirectly (super-learning). Overall, participants were able to learn when to respond and when to withhold their responding, although they did show an increased proportion of errors in the shift phase, suggesting that they found it difficult to respond to a cue that previously signalled the need to withhold responding. The correlational analysis suggested that psychopathic traits were negatively related to this change in errors, implying that individuals with higher levels of psychopathic traits found it easier to reverse their responding to the test cue. However, neither SCI nor FD predicted this difference in the mediation analysis. Participants' ratings showed that they were also able to learn about an inhibitory cue, and to flexibly update this learning. This ability was unrelated to psychopathic traits in either the correlational or mediation analyses.

The results appeared to show that the ability to use the inhibitor to guide subsequent learning was dependent upon sex, with only female participants showing significant super-learning. This suggests that only the female participants acquired inhibition in the acquisition phase and so were able to pass this second-order test of inhibitory learning. But, the imbalance in the number of male (n = 20) and female (n = 58) participants means that interpretation of this finding is difficult. Future experiments must use equal numbers of male and female participants to ensure that any

sex differences are not due to power, and then it will be possible to interpret any differences in learning associated with sex.

2.5. Conclusions

Experiment 1 established the basic principles of a novel method designed to assess associative learning in psychopathy. The results showed that explicit ratings can be used to directly and indirectly measure learning, and that this can be assessed separately to motor responses. It does not appear that psychopathic traits are predictive of general learning impairments. However, it remains to be seen how factors known to influence learning, such as attention (Mackintosh & Holgate, 1968) and outcome processing (Le Pelley, Mitchell, & Johnson, 2013) affect learning in relation to psychopathic traits, something that will be explored in Chapter 3.

CHAPTER 3

Testing the Conditions of Learning Differences in Psychopathy

(Experiments 2 – 5)

3.1. Introduction

As discussed in Chapter 1 (1.3.), theories of psychopathy typically propose either altered attention or reduced outcome processing are responsible for the differences seen in psychopathic individuals' learning and behaviour. I argued (1.5.) that these psychopathy accounts might be considered through their impact on the associability parameter (alpha, α) outlined in Mackintosh's (1975) Theory of Attention. For example, outcome processing accounts of psychopathy (e.g., Blair, 2005) suggest that psychopathic individuals' ability to process affective outcome information is impaired, which would mean that the affective value of a cue could not be used to facilitate learning about the cue like it would in non-psychopathic individuals (Le Pelley et al., 2013). Likewise, attentional accounts (e.g., Gorenstein & Newman, 1980) suggest that psychopathic individuals have overly selective attention biased in favour of cues central to the current goal focus, which may influence the perceived salience of a cue, and so skew the judgement of its relative validity, thereby biasing learning towards some cues and not others. Both these effects would influence the associability of a cue, something Mackintosh argued was responsible for determining how easily a cue is learnt about.

However, attention and outcome processing have also been shown more generally to affect learning, irrespective of psychopathic traits (e.g., Le Pelley et al., 2014, 2013).

As Experiment 1 suggested that psychopathic traits are not predictive of general learning impairments, the next step was to test whether the known effects of stimulus salience and outcome processing on learning are present regardless of the level of psychopathic traits; or, whether the reported disruption of these processes in psychopathy means that attention and outcome processing differentially affect psychopathic learning. This was tested in relation to cues' relative predictiveness, so how uniquely good the cue was at predicting the presence or absence of an outcome.

Specifically, the influence of contingency salience and predictiveness was tested by varying the magnitude of change in associative strength that occurred between the acquisition and shift phases. The present experiments were designed to test the relationship between psychopathic traits in community samples (students and non-students) and the acquisition and shift of excitatory and inhibitory associations. As in Experiment 1, these processes were assessed directly through participants' responses and ratings, and indirectly through the impact of this learning on subsequent association formation (e.g., super-learning; B. A. Williams & McDevitt, 2002).

3.2. General Method

3.2.1. Psychopathic Personality Inventory.

The details of the Psychopathic Personality Inventory (Revised; PPI-R; Lilienfeld & Widows, 2005) are the same as in Experiment 1, and will be the same throughout the remaining experiments. A brief summary is provided: this self-report questionnaire assessed psychopathic traits using 154 items, sixty-four of which were reverse scored. Each item was rated on a scale of: false, mostly false, mostly true or true. *T*-transformations were carried out on the Total PPI-R scores and the score for each of the three factors (Self-Centred Impulsivity, SCI; Fearless Dominance, FD; and,

Coldheartedness, C) on the basis of sample type (community vs. forensic), sex and age group of the participant. The internal consistency of the scale is reported separately for each experiment, as are participants' means and standard deviations for the Total PPI-R *t*-score, and the SCI, FD and C *t*-scores.

3.2.2 Novel Shift Go/NoGo Task

As shown in Figure 5, a new task design was used to facilitate the exploration of salience and outcome processing mechanisms, whilst separately testing stimulus-response and stimulus-outcome associations.

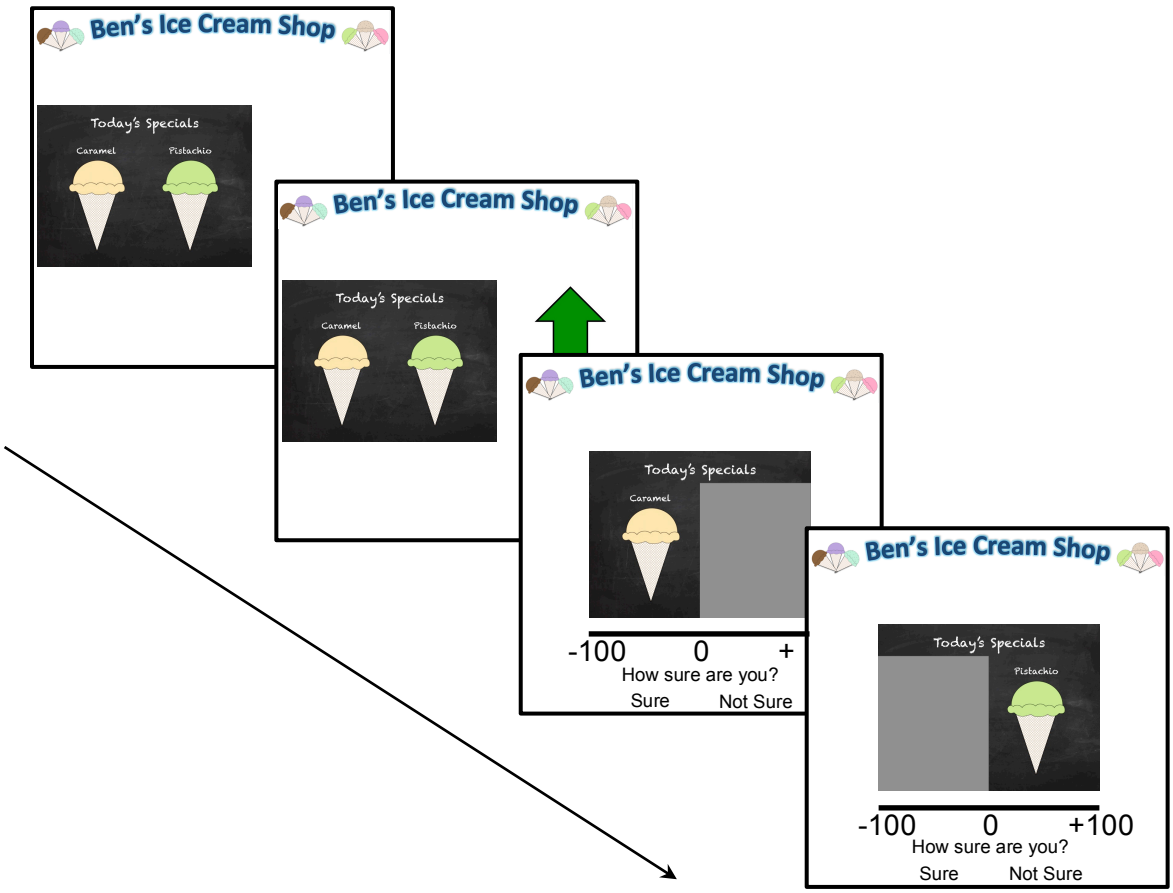


Figure 5. Experiments 2-6 example trial sequence for an excitatory compound cue.

The learning design of each experiment differed slightly to facilitate testing of specific experimental hypotheses, but the same general method was used in each study. This shared method will be outlined below, and the learning design for each experiment will then be shown in the relevant experiment method section. As before, the task required participants to learn associations between stimuli, responses and outcomes. Following the presentation of a stimulus (either an element or compound), participants were required to respond to a green arrow to indicate that the stimulus shown increased sales, and to withhold their responding to a red arrow to indicate that the stimulus shown decreased sales. This new procedure allowed separate testing of stimulus-response associations (S-R), which were assessed through the accuracy of these responses, and stimulus-outcome associations (S-O), which were assessed through stimulus ratings. Unlike in the previous design, participants rated the cues on every trial. Participants were also required to make a binary confidence judgement for each cue rating they made, indicating whether they were 'sure' (assigned a value of 1) or 'not sure' (assigned a value of 0) about their rating. However, none of the confidence ratings for the test cues changed significantly in any of the experiments, nor did psychopathy predict confidence ratings. As such, the confidence ratings will not be reported or further discussed for any of the experiments.

The stimuli were consistent with the selected cover story of monitoring ice cream sales, meaning each cue was represented by a flavour (both the colour of the 'scoop' and the name were shown). As shown in Figure 5, each trial began with the presentation of a stimulus, which was either an element (one ice cream flavour) or a compound (two ice cream flavours) on the screen for 500ms. After 500ms an arrow appeared alongside the stimulus for a maximum of 1000ms, though the cues were removed and the trial progressed if participants responded (by pressing the 'b' key). The arrow was either: a

green arrow pointing up accompanied by a high tone, signalling the need to respond; or, a red arrow pointing down accompanied by a low tone, signalling the need to withhold the response. Red and green arrows were used so as to make the task more closely resemble standard go/nogo assessments in which the go and nogo stimuli are distinguishable on the basis of a single feature, such as colour or letter type (Falkenstein, Hoormann, & Hohnsbein, 1999; Menon, Adleman, White, Glover, & Reiss, 2001). The red and green arrows acted as the outcome in the association with the stimulus (ice cream), signalling whether ice cream sales 'increased' or 'decreased'. The use of cues distinguishable on the basis of a single element was seen as preferable to having participants read sentences, such as they had to in Experiment 1. The high and low tones were included so as to support this easy discrimination between the stimuli. As discussed in Experiment 1 (2.2.3.), this type of bidirectional outcome task has previously been shown to produce conditioned inhibition and summation comparable to that seen when using a unidirectional outcome (Baetu & Baker, 2010).

The outcome (arrow) was probabilistic, occurring in one direction on 87.5% of trials for a given cue. Each stimulus was presented eight times in either the acquisition phase or the shift phase, meaning that a stimulus was reinforced in one direction on 7/8 trials, and in the opposite direction on 1/8 trials. After a response was made, or the maximum time reached, participants were shown the stimulus and asked to rate it on a +1/-1 increment scale from -100 (inhibitory) to +100 (excitatory). Where a compound stimulus had been seen, one side of the image was covered over and participants rated the remaining image before this was reversed and they rated the other part of the compound.

Therefore, as in Experiment 1, participants' performance on this novel go/nogo task was tested through their: *initial cue learning*, the difference between excitatory and inhibitory cues in the acquisition phase; *shift errors*, the difference in response accuracy

to a cue following contingency changes between the acquisition phase and the shift phase; *shift ratings*, the difference in ratings to a cue following contingency changes between the acquisition phase and the shift phase; and, *super-learning*, the difference in ratings between the super-learnt cue and a comparison cue in the shift phase. The last three variables were assessed in relation to both excitatory and inhibitory cues.

As in Experiment 1, two stimulus counterbalance conditions were included to ensure that any effects were not due to the flavour allocated to a stimulus. The design of the counterbalance is the same as that listed in Experiment 1 (pg. 50), with the addition that the negative super-learnt cue (W) was counterbalanced with its comparison cue (Z) in the same manner as the positive super-learning cues Y and S that were detailed in Experiment 1.

Non-colour-blind participants were recruited from the community in a mixed sample of students and non-students, and were given either course credit or payment. As in Experiment 1, all the studies conducted here were carried out under ethical approval granted by the University of Oxford Research Ethics Committee reference number MSD-IDREC-C2-2013-008.

3.2.3. Statistical Analysis

As in Experiment 1, initial acquisition of excitatory and inhibitory associations was tested by comparing ratings to the inhibitory test cue (A) and the excitatory test cue (B) in the acquisition phase, and also to a neutral test cue (C) in Experiment 5. In Experiment 2, the retardation of shift learning to the inhibitor was also tested by comparing the inhibitory cue (A) in the shift phase to a novel excitatory cue also presented in the shift phase (S). These tests were conducted in the same manner as in Experiment 1, by using mixed model analyses of variance, with the within-subjects factor of cue type, and the between-subjects factors of sex and stimulus counterbalance conditions. As outlined in

Experiment 1 (2.2.5.), stimulus counterbalance condition was included to ensure that any effects were not due to certain ice cream flavours being more easily learnt about in association with certain outcomes, and sex was included given previous demonstrations of sex differences in relation to both the conceptualisation of psychopathy, and how psychopathy relates to external variables (Cale & Lilienfeld, 2002; Z. Lee & Salekin, 2010; Lilienfeld & Hess, 2001). Subsequent design changes meant that the retardation test was not possible in Experiments 3-5, and so assessment of inhibitory learning was done directly by testing the shift in learning to the inhibitor itself, and indirectly against an excitatory background by testing its ability to influence learning about other cues (e.g., super-learning).

Having established these basic acquisition effects, overall shift and super-learning effects were tested. In each experiment, the dependent measures for both the excitatory and inhibitory cues were: shift errors; shift ratings; and, super-learning ratings. As in Experiment 1, the shift error variables were calculated by subtracting the errors made in the shift phase from those made in the acquisition phase, meaning a negative score indicated that participants made more errors in the shift phase than in the acquisition phase; the shift ratings were calculated by subtracting the acquisition phase ratings from the shift phase ratings, meaning that for the inhibitory cue (A) a positive rating indicated greater shift, but for the excitatory cue (B) a negative rating indicated great shift; and, for the super-learning cues the rating of the comparison cue was subtracted from that of the super-learnt cue, meaning that for positive super-learning a positive value indicated greater super-learning, whilst for negative super-learning a negative value indicated greater super-learning. Mixed model analyses of variance (ANOVA) were used to assess the shift go/nogo errors and shift ratings in relation to the within-subjects variable of phase (acquisition vs. shift), with sex and flavour counterbalancing conditions also

included as between-subjects variables. A mixed model ANOVA was also used to assess super-learning ratings with respect to the within-subjects variable of cue type (super-learnt vs. comparison), again with sex and flavour counterbalancing conditions included as between-subjects variables. *T*-tests were only conducted following significant results of ANOVA and were corrected using Bonferroni.

The relationship between the excitatory and inhibitory dependent variables and psychopathy traits were also assessed. Zero-order correlations were calculated between each of the dependent variables, and the overall psychopathy score (total PPI-R score), and each of the PPI-R factors (SCI, FD and C). In addition, multiple mediation analyses with mediators operating in serial were conducted to test whether either FD or SCI predicted super-learning, and whether this relationship was mediated by shift errors or shift ratings to the relevant cues. As explained in Chapter 2 (2.2.5.) shift errors were included as a mediator variable because responses can alter learning about contingencies (Matute, 1996); and shift ratings were included as a mediator variable because the super-learning was dependent on prior learning about the shift cue. These relationships were tested using a non-parametric bootstrapping procedure (PROCESS; Preacher & Hayes, 2008) which estimates the indirect effects of multiple mediator variables, and as in Experiment 1 was based on 1000 bootstrap resamples with 95% bias corrected confidence intervals. Sex and flavour counterbalancing conditions were controlled for in the models.

As in Experiment 1, an alpha level of .05 was used throughout and 95% confidence intervals were calculated for all analyses of variance, *t*-tests and mediation analyses where $p < .05$. Confidence intervals are reported to two decimal places and so are sometimes shown as .00, even though this may represent a value between zero and .005. Where .00 represents a value that is equal to or less than zero, this will be explicitly

stated and the implications of this on the interpretation of any difference between means will be discussed.

3.3. Experiment 2

3.3.1. Introduction

Having established in Experiment 1 that a go/nogo task could be used to train associative learning, this experiment sought to test the acquisition and flexible shifting of associations. As in Experiment 1, there was a large magnitude of change in associative strength of the test cue(s) between the acquisition and shift phases so as to offer a clear test of participants' ability to monitor contingencies. It was predicted that, like in Experiment 1, participants' psychopathic traits would not be related to their ability to learn or reverse their learning. Building on the results of Experiment 1, this study looked at both excitatory and inhibitory learning to test the symmetry of these processes in relation to psychopathic traits. This balance between excitatory and inhibitory processes is important because any apparent inhibitory problems may reflect poor inhibition, over-developed excitation or a combination of the two. As such, the balance of excitatory and inhibitory processes was tested through the direct assessment of shift errors and shift ratings to the excitatory and inhibitory cues, and through the indirect assessment of the cues' ability to influence learning about novel cues.

Specifically, the experiment tested whether participants showed super-learning: the facilitation of enhanced learning about a novel cue when it is trained in compound with a cue previously established to predict the opposite outcome (B. A. Williams & McDevitt, 2002). The inhibitor (A) was predicted to facilitate 'positive super-learning' or enhanced excitatory learning about a novel cue (Y) with which it was positively reinforced (AY+); whilst the excitor (B) was predicted to facilitate 'negative super-learning' or enhanced inhibitory learning about a cue (W) with which it was negatively reinforced (BW-). There is little previous evidence of such negative super-learning. So, in addition to providing an indirect test of the balance in excitatory and inhibitory learning in

psychopathy, by testing this procedure I sought to add to previous work regarding the symmetry of excitatory and inhibitory processes (Rescorla, 1969).

In addition to the go/nogo task this experiment included an alternative, independent assessment of participants' inhibition: a stop-signal task (SST; Verbruggen, Logan, & Stevens, 2008). Whilst the results of Experiment 1 contradicted some of the previous findings in the literature that have shown increased impulsivity in relation to psychopathic traits, it is noted that a number of these studies used an imbalance in go/nogo trials or pre-exposure to go trials only (Lapierre et al., 1995; Newman et al., 1990). As was discussed in Chapter 1 (1.6.), such an imbalance in trial type can affect the attentional demands of responding or withholding a response. Therefore, by inducing such a bias a go/nogo response inhibition task may become more like a stop-signal response cancellation task. Given the present go/nogo design used a balanced number of go (excitatory) and nogo (inhibitory) trials, a stop-signal task was included to test whether cancellation abilities are affected by psychopathic traits.

3.3.2. Method

3.3.2.1. Participants

Thirty-four (17 male, 17 female; age 18-25 years) individuals participated under the recruitment and ethical procedures outlined in the General Method section of this chapter (3.2.2). The number of participants was determined by a power analysis conducted on the basis of pilot data negative super-learning ratings using this new task. The results of Experiment 1 were not used in the power analysis as the rating scales were different and negative super-learning was not tested in Experiment 1, and so pilot data from the modified task was deemed more appropriate. On the basis of the pilot data super-learnt and comparison cue ratings (μ_1 = -56.97; μ_2 = -40.13; σ = 24.53) a power

analysis indicated that 34 participants were needed to have 80% power for detecting an effect when employing the traditional .05 criterion of statistical significance.

3.3.2.2. Design

In the go/nogo task shift learning (responses and ratings) and super-learning were assessed using the 2 x 2 x 2 mixed model design outlined in the Statistical Analysis section (3.2.3). For the stop-signal task, a between subjects' design was used to compare participants' Stop-Signal Reaction Time (SSRT) in relation to their psychopathic traits. This was done using a simultaneous multiple linear regression analysis.

3.3.2.3. Materials

Psychopathic Personality Inventory (Revised; PPI-R; Lilienfeld & Widows, 2005). In the present sample, the internal consistency of the PPI-R was good (α = .91). The means and standard deviations of the participants are shown in Table 4.

Table 4

Experiment 2 Participant Psychopathy *T*-Scores

	M	*SD*
Total PPI-R	49.85	8.50
Self-Centred Impulsivity	52.12	7.99
Fearless Dominance	48.06	9.07
Coldheartedness	46.58	9.51

Note: N = 34. The mean *t*-score is 50.

Shift Go/NoGo Task. The general method is outlined in section 3.2.2. of this chapter. As shown in Table 5, the presentation of excitatory and inhibitory cues was balanced within each phase to ensure that any shift effects were not due to participants

having only received inhibitory cues in the acquisition phase and excitatory cues in the shift phase or vice versa. Included in the design were two test cues (A and B) that had their associations shifted between the two phases. A was trained as inhibitory in the acquisition phase (AX-, X+; Rescorla, 1969), and was then trained as excitatory in the shift phase (AY+). The total response errors to A in the acquisition (Aq) and shift (S) phases were compared to determine whether shift learning occurred ($A_{Aq} = A_S$) or not ($A_{Aq} < A_S$). Likewise the average ratings of A in the acquisition and shift phases were compared to determine whether shift learning occurred ($A_{Aq} < A_S$) or not ($A_{Aq} = A_S$).

Table 5

Experiment 2 Design

Acquisition Phase	**Shift Phase**
AX-, X+	AY+
BV+, V-	BW-
	CZ-, C+
	DS+, D-

Note: Overall outcome direction is shown as excitatory (+) or inhibitory (-). A and B are the shift test cues, and Y and W are the super-learning test cues. All other cues are included to counterbalance the training of A, B W and Y and inhibitory-excitatory exposure.

The second-order, indirect test of A was to assess whether it was used to facilitate enhanced excitatory learning about a cue with which it was trained in compound in the

shift phase (Y). As outlined in Chapter 1 (1.7.), such super-learning occurs when the disparity between the expected ($\Sigma V_{AY} = -1$) and the actual ($\lambda = +1$) outcome is large (D. A. Williams et al., 2004). This positive super-learnt cue was compared to a novel cue (S) also trained in compound and paired with the outcome (DS+) to see if super-learning occurred (Y > S) or not (Y = S).

Cue B was paired with the presence of the outcome in the acquisition phase (BV+), and the absence of the outcome in the shift phase (BW-). As with A, response errors to B in the acquisition and shift phases were compared to determine whether shift learning occurred ($B_{Aq} = B_S$) or not ($B_{Aq} < B_S$); and also average ratings of B in the acquisition and shift phases were compared to determine whether shift learning occurred ($B_{Aq} > B_S$) or not ($B_{Aq} = B_S$). The second-order, indirect test of the excitatory strength of B was to see if it facilitated enhanced inhibitory learning about a novel cue with which is was negatively reinforced in compound (BW-). Such negative super-learning was compared to a novel cue (Z) also trained in compound to predict the absence of the outcome (CZ-) to test if negative super-learning occurred (W < Z) or not (W = Z).

Stop-Signal Task (SST; Verbruggen et al., 2008). This was used to assess response inhibition. Participants were asked to differentially respond as quickly as possible to two randomly presented shapes (square or diamond) by pressing a corresponding computer key ("c" or "m"). They were told that either of the shapes could be followed by the stop-signal (the emboldening of the shape), and that when this happened they should press neither key. On 25% of trials participants saw the stop-signal. The two trial types were randomly presented. A fixation-cross appeared for 250ms between trials. Stimuli remained on screen until the participant pressed either key or until 1250ms had elapsed. The task consisted of 32 practice trials followed by three blocks of 64 trials. Participants were given feedback on their performance at the end of each block.

The Stop-Signal Distance (SSD: time between presentation of the shape and the stop-signal) varied depending on participants' performance. Successful inhibition increased the SSD by 50ms, whilst failure to inhibit reduced it by 50ms. This enabled participants' mean SSD across trials to be calculated. The SSD was varied to keep participants' performance at 50% correct. Participants' Stop-Signal Reaction Time (SSRT) was calculated as the SSRT-integration: the calculated point at which the stop process finishes minus the mean SSD. The calculated point at which the stop process finished is determined by multiplying the probability of participants responding given the stop-signal appeared by their final (*n*th) RT. The SSRT-integration was used instead of the SSRT-mean (in which the mean SSD is subtracted from the mean reaction time). This is because recent research on the accuracy of these two methods in estimating stopping times showed that the SSRT-mean method consistently overestimates the SSRT, and thus leads to spurious inhibitory effects (Verbruggen, Chambers, & Logan, 2013).

3.3.2.4. Procedure

Having given their consent, participants first completed the go/nogo task, then the SST (Verbruggen et al., 2008), then the PPI-R (S. O. Lilienfeld & Widows, 2005), before being debriefed.

3.3.3. Results

In the go/nogo task participants made more errors to the excitatory cue in the acquisition phase than in the shift phase, but shift errors for both the excitatory and inhibitory cues were unrelated to psychopathic traits. As shown in Figure 7, participants showed a significant shift in ratings to the excitatory and inhibitory test cues, and neither FD nor SCI predicted this shift in association. There was significant evidence of negative

super-learning, but not positive super-learning. In the stop-signal task participants SSRT-integration was not significantly related to either FD or SCI.

3.3.3.1. Overall Learning Effects

Initial Cue Learning (Acquisition Phase). As shown in the upper panel of Figure 6 participants rated the inhibitory cue (A; M = -47.35, SD = 16.63) as significantly more negative than the excitatory cue (B; M = 51.03, SD = 18.48), $F(1,29) = 220.80$, $p < .001$, $\eta_p^2 = .88$, 95% CI (.78, .92). No interactions were significant, all Fs < .67, all ps > .42.

Initial Cue Learning (Shift Phase). As shown in the lower panel of Figure 6 participants showed a retardation effect, rating the previously trained inhibitory cue, which is trained as excitatory in the shift phase (A; M = -20.39, SD = 37.90) as significantly more negative than a novel excitatory cue (S; M = 45.49, SD = 23.04), $F(1,29) = 58.80$, $p < .001$, $\eta_p^2 = .67$, 95% CI (.43, .78). No interactions were significant, all Fs < .52, all ps > .31.

Shift Errors. Participants did not make a significantly different number of errors to the inhibitory cue (A) in the acquisition phase (M = .61, SD = .83) and the shift phase (M = .42, SD = 1.28), all Fs < 1.40, all ps > .24. Participants made significantly more errors to the excitatory cue (B) in the acquisition phase (M = .91, SD = 1.59) than in the shift phase (M = .15, SD = .36), $F(1,29) = 4.82$, $p = .036$, $\eta_p^2 = .14$, 95% CI (.00, .36). No interactions were significant, all Fs < .46, all ps > .50.

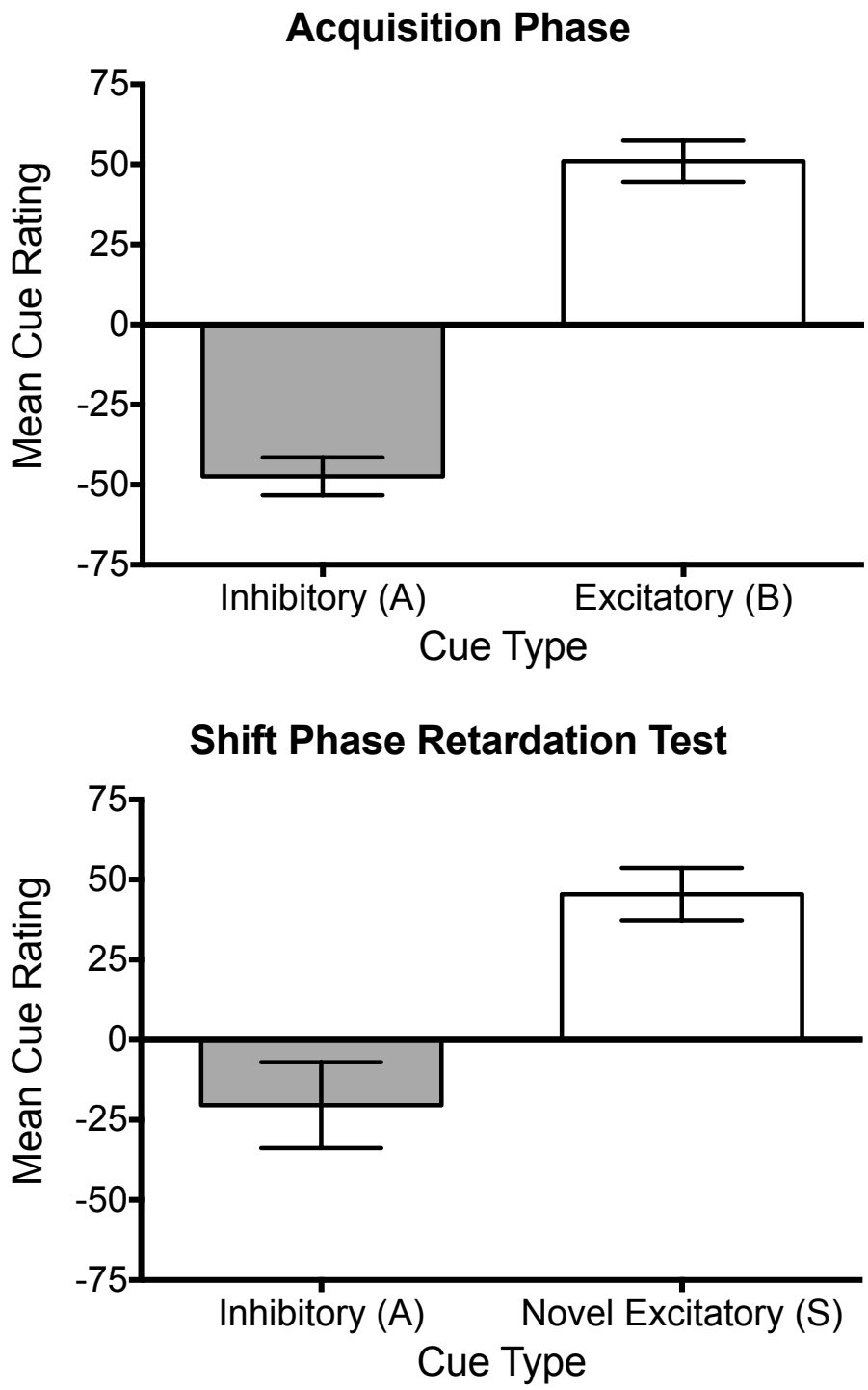


Figure 6. The upper panel shows that participants rated the inhibitory (A) and the excitatory (B) cues as significantly different in the acquisition phase. The lower panel shows a retardation effect, whereby the inhibitory cue (A), which is subsequently trained as an excitor in the shift phase is rated significantly lower than a novel cue (S) also trained as an excitor in the shift phase. Error bars represent 95% confidence intervals.

Shift Ratings. As shown in Figure 7, overall participants shifted their learning to the previously predictive inhibitor (A; acquisition: M = -47.35, SD = 16.63; shift: M = -20.39, SD = 37.90), $F(1,32) = 16.69$, $p < .001$, $\eta_p^2 = .34$, 95% CI (.09, .54). No interactions were significant, all Fs < 2.38, all ps > .13. Also shown in Figure 7, overall participants shifted their learning to the previously predictive excitor (B; acquisition: M = 51.03, SD = 18.48; shift: M = 31.44, SD = 34.43), $F(1,32) = 9.96$, $p = .004$, $\eta_p^2 = .26$, 95% CI (03, .45). No interactions were significant, all Fs < .30, all ps > .69.

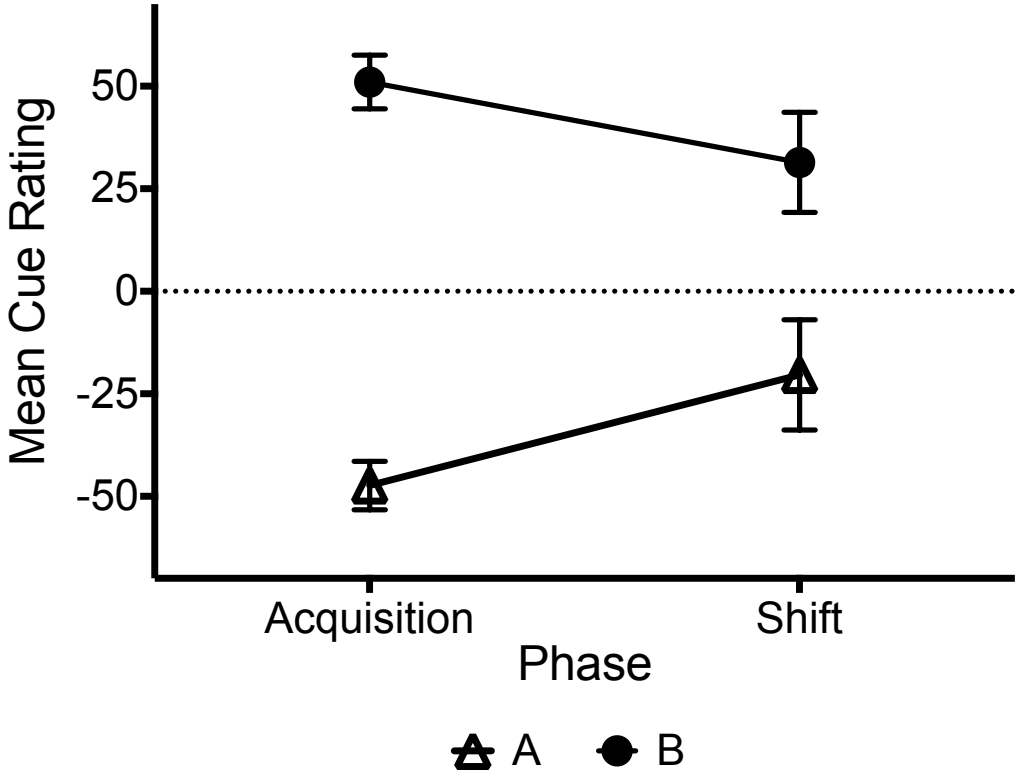


Figure 7. Participants showed significant shift learning to the inhibitory (A) and the excitatory (B) test cues. Error bars represent the 95% confidence intervals.

Super-Learning. Overall, participants did not show enhanced learning to Y (M = 48.92, SD = 26.53) as compared to S (M = 45.49, SD = 23.04), so-called positive super-learning, all Fs < .31, all ps > .58. However, participants did show enhanced learning to W (M = -54.89, SD = 22.71) as compared to Z (M = -43.30, SD = 23.73), so-called negative super-learning, $F(1,32) = 5.82$, $p = .022$, $\eta_p^2 = .15$, 95% CI (.01, .37). No interactions were significant, all Fs < 1.50, all ps > .23.

3.3.3.2. Relationship Between Psychopathy and Dependent Variables

The zero-order correlations between overall psychopathy score, the factors of SCI, FD and C, and the dependent variables are shown in Table 6. The dependent variables were not significantly correlated with psychopathy traits.

The mediation analysis showed positive super-learning was not directly or indirectly (through shift errors and/or shift ratings) predicted by FD [all βs < .23, all 95% CIs (< -.07, > .09)] or SCI [all βs < .42, all 95% CIs (< -.22, > .09)]. Likewise negative super-learning was not directly or indirectly predicted by FD [all βs = .66, all 95% CIs (< -.11, > .04)] or SCI [all βs < .20, all 95% CIs (< -.15, > .06)]. However, shift learning to the inhibitory cue (A) did significantly predict the extent of positive super-learning in both the FD [β < .61, 95% CI (.30, .92), $t(33) = 3.96$, $p = .001$, $\eta_p^2 = .32$, 95% CI (.08, .52)] and SCI models [β = .61, 95% CI (.30, .92), $t(33) = 4.02$, $p < .001$, $\eta_p^2 = .33$, 95% CI (.09, .52)]. Likewise, shift learning about the excitatory cue (B) significantly predicted the extent of negative super-learning in both the FD [β = .48, 95% CI (.18, .71), $t(33) = 3.88$, $p = .001$, $\eta_p^2 = .31$, 95% CI (.08, .51)] and SCI models [β = .45, 95% CI (.18, .71), $t(33) = 3.49$, $p = .002$, $\eta_p^2 = .27$, 95% CI (.05, .47)].

Table 6

Experiment 2 Zero-Order Correlations

	Total PPI-R	SCI	FD	C
Total PPI-R	-			
SCI	**.752*** **	-		
FD	**.746*** **	.196	-	
C	.331	.024	.162	-
A Shift Errors	-.046	-.164	.103	.009
B Shift Errors	.099	.087	.053	.033
A Shift Ratings	.023	-.040	.097	-.047
B Shift Ratings	-.079	.072	-.130	-.235
Positive SL	.083	.060	.024	.101
Negative SL	.125	.076	.135	-.023
SSRT	.161	.077	.043	.076

Note: Total PPI-R, SCI, FD and C represent the *t*-scores for each variable. Positive SL = Positive Super-Learning Ratings; Negative SL = Negative Super-Learning Ratings; SSRT = Stop-Signal Reaction Time.

*** $p < .001$; ** $p < .01$; * $p < .05$.

Stop-Signal Reaction Times. The multiple linear regression analysis showed that the psychopathy factors (FD and SCI) did not significantly predict participants' SSRTs, $R^2_{adjusted} = -.046$, $F(4,32) = .65$, $p = .596$, all βs < .20, all ps > .30.

3.3.4. Discussion

Unlike in Experiment 1, participants did not appear to find it difficult to alter their responding to the test cues, and in fact participants made more errors to the excitatory cue in the acquisition phase than they did in the shift phase. Psychopathic traits were not predictive of a difference in errors across the phases for either the excitatory or inhibitory cue. Consistent with Experiment 1, the absence of behavioural differences in relation to psychopathy suggests that individuals are capable of shifting their learning on the basis of a predictive rule, and are able to use this to guide their responses. This was supported by the absence of a relationship between psychopathic traits and SSRT, providing additional evidence independent of the go/nogo task developed here.

It is worth noting that the overall number of errors made by participants was low, so it is possible that the task was too easy and failed to distinguish participants' abilities. However, previous studies have often used a greater number of go trials than nogo trials in order to induce a response tendency in participants (Lapierre et al., 1995; Munro et al., 2007b). This method biases participants into responding, which, whilst interesting, does not assess their ability to inhibit a response that is equally as predominant as not responding. Under these biased conditions, participants will require greater attentional allocation to nogo trials in order to inhibit their response (Kiehl et al., 2000). Given the suggestion that attentional allocation in psychopathic individuals is altered (Hiatt, Schmitt, & Newman, 2004; Newman & Baskin-Sommers, 2011; Wolf et al., 2012), findings from studies using an unequal number of go and nogo trials may reflect loading on to mechanisms (i.e., attentional allocation) known to be altered in psychopathic individuals, producing a 'false-positive'.

The present behavioural findings are consistent with previous work that also used equal numbers of go and nogo trials (Kiehl et al., 2000). For instance, Kiehl and

colleagues found no difference in the errors made by psychopathic and non-psychopathic forensic samples. Indeed, the low number of errors reported here is consistent with Kiehl et al.'s study in which they showed that psychopathic and non-psychopathic participants made a mean number of approximately 4.96 errors across 540 trials. This number of errors (approximately 1%) is in fact lower than that recorded here to the test cues (approximately 6%). As such, participants' overall high levels of accuracy are consistent with previous research, and do not appear to simply suggest a ceiling effect.

Participants' ability to monitor contingencies and so shift their learning on the basis of predictive information was also reflected in their cue ratings, which changed significantly between the acquisition and shift phases for both the excitatory and inhibitory cues. Like the changes in responses, neither SCI nor FD predicted these significant shifts. This symmetry of shift learning to both the previously trained inhibitory cue and the previously trained excitatory cue suggests that there is a balance in participants' ability to learn from predictive excitatory and inhibitory information, as determined through direct assessment of the cue. However, this symmetry was not seen in the second-order effects, as measured by the ability of the test cue to facilitate enhanced (super) learning about another cue. No overall positive super-learning effect was seen, suggesting that participants failed this second-order test of inhibition. This was not related to psychopathic traits. In contrast, negative super-learning was shown, and I believe that this is the first demonstration of such an effect. This asymmetry suggests a difference in the way in which the inhibitor (necessary to facilitate positive super-learning) was learnt about, and implies that in the shift phase the inhibitory strength of test cue A extinguished more rapidly than the excitatory strength of test cue B, meaning that A was unable to facilitate super-learning. Neither FD nor SCI predicted this difference.

Like Experiment 1, these results do not support an outcome processing account (e.g., IES; Blair, 2005). Outcome processing accounts such as the IES suggest that psychopaths have a difficulty in using goal-directed processes that are dependent on forming associations with the outcome. If an outcome processing account were true, there would have been evidence of generally impaired learning about the cues, particularly in the shift phase. No evidence of such a general learning impairment was seen. It could be argued that outcome-processing effects influenced positive super-learning. However, as this was not related to psychopathic traits, nor is it obvious why an outcome processing problem would disrupt only positive super-learning and not any of the other forms of learning tested here, these results do not appear to support an outcome processing account.

Overall the results of Experiment 2 are consistent with those of Experiment 1, showing that participants are able to monitor contingencies, and so learn and shift their learning about inhibitory cues. Experiment 2 extended this finding by demonstrating that this ability also encompasses excitatory cues, thereby suggesting that a modified go/nogo task can be used to train both excitatory and inhibitory associations. Having established these basic learning effects, this novel paradigm can now be used to test whether psychopathic traits are associated with differences in contingency sensitivity and how this affects excitatory and inhibitory learning. In doing so it will be possible to test the predictions made by attentional accounts and outcome processing accounts of psychopathy.

3.4. Experiment 3

3.4.1. Introduction

Experiments 1 and 2 demonstrated that psychopathic traits are not predictive of differences in learning and shifting learning on the basis of predictive information when the change in contingency is large (i.e., from inhibitory to excitatory). This study sought to build on these findings by looking at how participants update their learning when the stimulus-outcome contingencies are changed, but not completely reversed on the basis of weakly predictive information. As previously discussed, theories of associative learning show that the predictiveness of a cue has an affect on the way in which it is learnt about (Le Pelley & McLaren, 2003; Mackintosh, 1975; Pearce & Hall, 1980; Rescorla & Wagner, 1972), something that Mackintosh suggested affects the cue's associability or relative validity. As determining a cue's associability is reliant on processing all the cues present, in Chapter 1 (1.5.) I suggested that this process may be disrupted by either altered attention or altered outcome processing, offering a potential mechanism by which theories of psychopathic impulsivity may operate.

Therefore, Experiment 3 tested whether variations in cues' predictiveness differentially affected participants' learning with respect to their level of psychopathic traits, specifically their levels of SCI and FD. In order to facilitate this, I tested the impact of reducing the test cues' predictiveness on the direct assessment of levels of shift learning, and the in-direct assessment of the impact that these test cues had on learning about other stimuli with which they are paired.

3.4.2. Methods

3.4.2.1. Participants and Design

Forty-seven (24 male, 23 female; age M = 21.43 years, SD = 2.46) individuals

participated, using the recruitment and ethical procedures outlined in the General Methods section of this chapter (3.2.2.). A power analysis on the basis of the negative super-learning in Experiment 2 (μ_1 = -54.89; μ_2 = -43.30; σ = 22.58) indicated that 47 participants were needed to have 80% power for detecting an effect when employing the traditional .05 criterion of statistical significance. The design was the same as Experiment 2.

3.4.2.2. Materials

Psychopathic Personality Inventory (Revised; PPI-R; Lilienfeld & Widows, 2005). In the present sample, the internal consistency of the PPI-R was good (α = .91). The means and standard deviations for the participants are shown in Table 7.

Table 7

Experiment 3 Participant Psychopathy *T*-Scores

	M	*SD*
Total PPI-R	50.74	10.24
Self-Centred Impulsivity	52.28	9.28
Fearless Dominance	48.85	9.59
Coldheartedness	50.49	12.39

Note: N = 47. The mean *t*-score is 50.

Shift Go/NoGo Task. In this design participants were required to learn when to respond and when to inhibit their responses to twelve different stimuli that had positive (excitatory: B+, D+, X+, AY+, BV+, DS+) or negative (inhibitory: A-, C-, V-, AX-, BW-, CZ-) contingencies, as shown in Table 8. As before, two test cues (A and B) had their associations shifted between the acquisition and shift phase, however here the shift

was done of the basis of more inconsistent information than that used in Experiment 2. As in Experiment 2, A was trained as a predictive inhibitor in the acquisition phase, meaning it was predictive of the absence of the outcome (AX-, X+; Rescorla, 1969).

Table 8

Experiment 3 Design

Acquisition Phase	Shift Phase
AX-, X+	AY+, A-
BV+, V-	BW-, B+
	CZ-, C-
	DS+, D+

Note: Overall outcome direction is shown as excitatory (+) or inhibitory (-). A and B are the shift test cues, and Y and W are the super-learning test cues. All other cues are included to counterbalance the training of A, B Y and W, and inhibitory-excitatory exposure.

The design differed from Experiment 2 as in this experiment A was trained as a weakly predictive inhibitory cue in the shift phase. This meant that A had low associability and so learning about A in the shift phase was predicted to be weak (AY+, A-; Mackintosh, 1975). The response errors to, and average ratings of A in the acquisition and shift phases were compared to determine whether shift learning occurred or not on the basis of weakly predictive information.

As before B was trained as a predictive excitor in the acquisition phase, meaning it was predictive of the presence of the outcome (BV+). B was then trained as a weakly predictive excitatory cue in the shift phase, meaning that, like A, it had low associability and so learning about B in the shift phase was predicted to be weak (BW-, B+; Mackintosh, 1975). As with A, total response errors to and average ratings of B in the acquisition and shift phases were compared to determine whether shift learning occurred or not on the basis of weakly predictive information. As in Experiment 2, test cue A was predicted to facilitate enhanced excitatory learning about cue Y, as compared to cue S; and test cue B was predicted to facilitate enhanced inhibitory learning in cue W, as compared to cue Z.

3.4.2.3. Procedure. Having given their consent, participants first completed the go/nogo task, then the PPI-R (Lilienfeld & Widows, 2005) before being debriefed. This procedure is used in all remaining studies.

3.4.3. Results

Consistent with Experiment 2, participants made more errors to the excitatory test cue in the acquisition phase than they did in the shift phase, but their errors to the inhibitory test cue did not significantly change across phases. Neither FD nor SCI predicted these response patterns. As shown in Figure 9, and consistent with Experiment 2, participants showed significant shift learning to both the excitatory and the inhibitory cue. Participants also showed significant positive and negative super-learning based on the inhibitory and excitatory test cues respectively, which is shown in Figure 10. Both FD and SCI predicted increased shift learning and reduced super-learning based on the inhibitory cue, but not the excitatory cue. Shift learning to the excitatory cue was

significantly predicted by sex, with females showing greater shift learning, and shift learning to the excitatory cue predicted reduced negative super-learning.

3.4.3.1. Overall Learning Effects

Initial Cue Learning (Acquisition Phase). As shown in Figure 8 participants rated the inhibitory cue (A; M = -48.89, SD = 20.40) as significantly more negative than the excitatory cue (B; M = 52.65, SD = 20.62), $F(1, 43) = 354.04$, $p < .001$, $\eta_p^2 = .89$, 95% CI (.82, .92). No interactions were significant, all Fs < 1.33, all ps > .25.

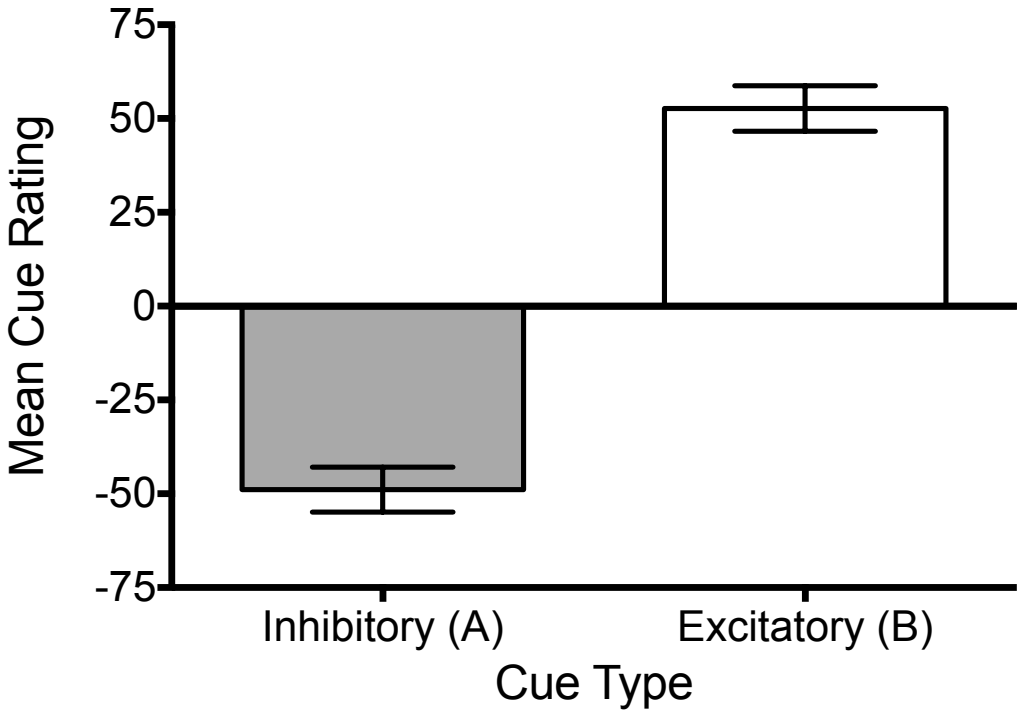


Figure 8. Participants rated the inhibitory cue (A) as significantly more negative than the excitatory cue (B) in the acquisition phase. Error bars represent 95% confidence intervals.

Shift Errors. Overall, participants did not make a significantly different number of errors to test cue A in the shift phase (M = .57, SD = .83) as compared to the acquisition

phase (M = .98, SD = 1.73), all Fs < 2.0, all ps > .16. In contrast, participants made significantly more errors to test cue B in the acquisition phase (M = 1.11, SD = 1.71) than in the shift phase (M = .11, SD = .31), $F(1,43) = 13.01$, $p = .001$, $\eta_p^2 = .23$, 95% CI (.05, .42). No interactions were significant, all Fs < 2.5, all ps > .11.

Shift Ratings. As shown in Figure 9, overall participants shifted their learning to the previously predictive inhibitor (A; acquisition: M = -48.89, SD = 20.40; shift: M = -38.43, SD = 23.45), producing a main effect of phase, $F(1,43) = 9.09$, $p = .004$, $\eta_p^2 = .18$, 95% CI (.02, .36). The interaction between phase and sex was also significant, $F(1,43) = 6.96$, $p = .012$, $\eta_p^2 = .14$, 95% CI (.00, .33). However, the 95% confidence intervals about the effect size crossed zero, suggesting that there may in fact have been no difference between male and female participants' shift learning to the inhibitor.

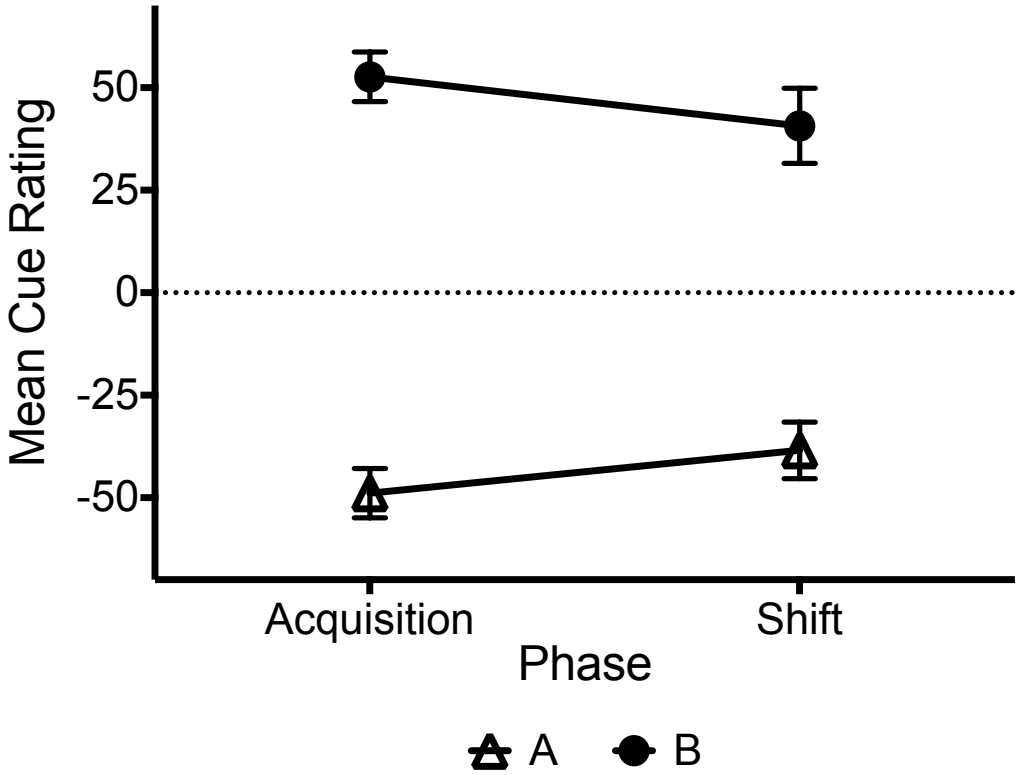


Figure 9. Participants showed significant shift learning to the inhibitory (A) and excitatory (B) test cues. Error bars represent 95% confidence intervals.

Also shown in Figure 9, participants shifted their learning to the previously predictive excitor (B; acquisition: $M = 52.65$, $SD = 20.62$; shift: $M = 40.71$, $SD = 31.12$), producing a significant main effect of phase, $F(1,43) = 7.15$, $p = .011$, $\eta_p^2 = .14$, 95% CI (.01, .33). The interaction between phase and sex was also significant, $F(1,43) = 8.88$, $p = .005$, $\eta_p^2 = .17$, 95% CI (.02, .36). *Post hoc t*-tests carried out to compare shift learning in males and females were significant for the females, $t(22) = 4.08$, $p < .001$, $\eta_p^2 = .43$, 95% CI (.11, .63), but not for the males, $t(23) = .03$, $p = .977$, $\eta_p^2 = .00$, 95% CI (.00, .00), indicating that only the female participants showed significant shift learning to the excitor.

Super-Learning. As shown in Figure 10, participants showed enhanced learning to Y ($M = 54.47$, $SD = 21.59$) as compared to S ($M = 43.52$, $SD = 23.16$), so-called positive super-learning, $F(1,43) = 9.83$, $p = .003$, $\eta_p^2 = .19$, 95% CI (.02, .37). No other interactions were significant, all Fs < 1.84, all ps > .18. Participants also showed enhanced learning to W ($M = -56.82$, $SD = 26.01$) as compared to Z ($M = -47.26$, $SD = 22.86$), so-called negative super-learning, $F(1,43) = 4.95$, $p = .031$, $\eta_p^2 = .10$, 95% CI (.00, .28). No other interactions were significant, all Fs < 2.75, all ps > .10.

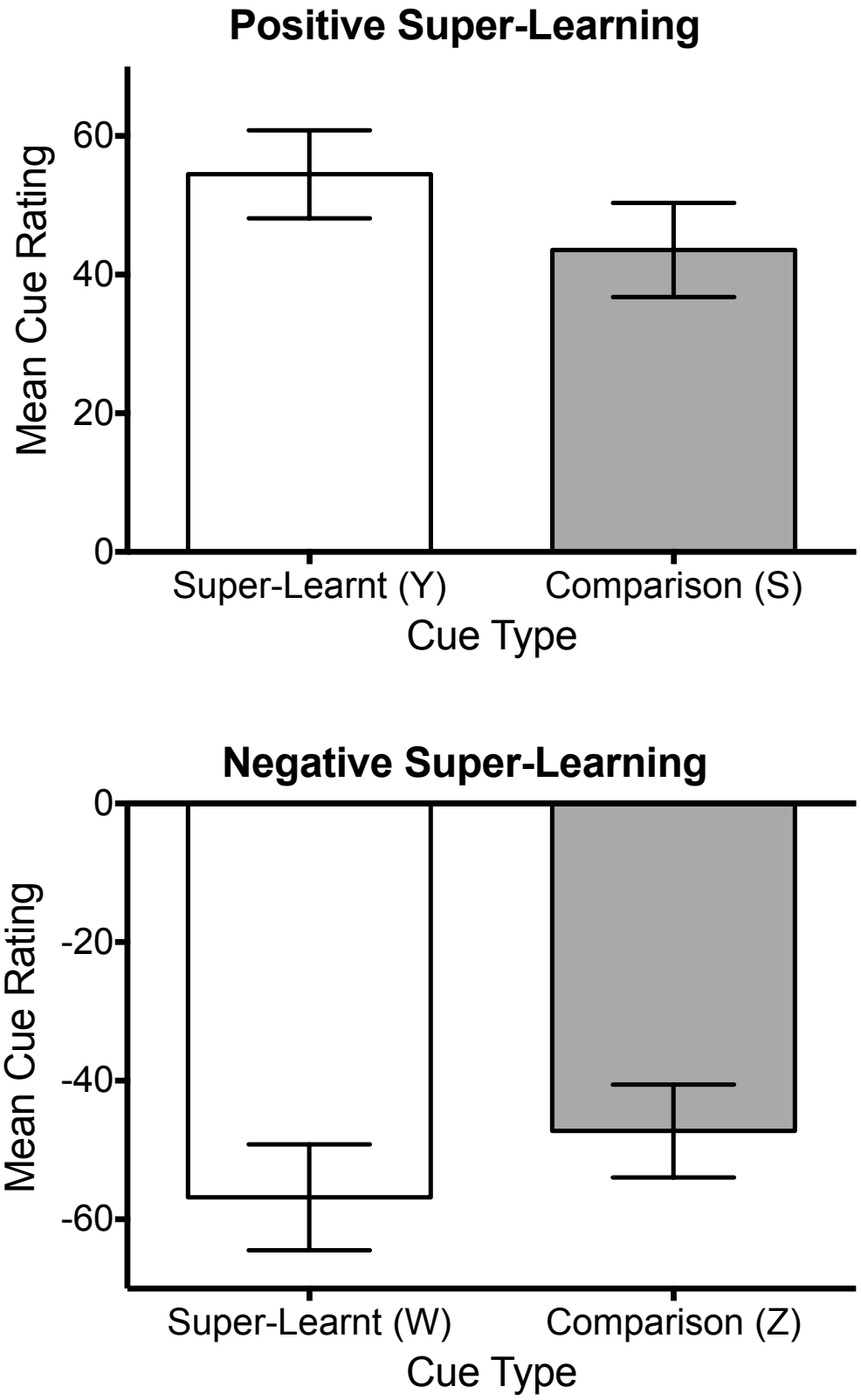


Figure 10. Participants showed significant super-learning to Y and W as compared with S and Z respectively. Error bars represent 95% confidence intervals.

3.4.3.2. Relationship Between Psychopathy and Dependent Variables

The zero-order correlations between overall psychopathy score, the factors of SCI, FD and C, and the dependent variables are shown in Table 9. As expected, each of the three PPI-R factors was strongly related to the overall PPI-R score. The shift in ratings to the inhibitory test cue (A) were significantly correlated with overall psychopathy traits (r = .40, p = .005), and with the factors of SCI (r = .31, p = .036) and FD (r = .38, p = .008). All other dependent variables were not significantly related to psychopathy.

Table 9

Experiment 3 Zero-Order Correlations

	Total PPI-R	**SCI**	**FD**	**C**
Total PPI-R	**-**			
SCI	**.794*****	-		
FD	**.703*****	.206	-	
C	**.362***	.199	.072	-
A Shift Errors	-.164	-.081	-.154	.009
B Shift Errors	.203	.138	.266	-.145
A Shift Ratings	**.400****	**.307***	**.384****	.021
B Shift Ratings	-.213	-.120	-.226	-.057
Positive SL	-021	.066	-.071	-.037
Negative SL	-.090	.150	.286	-.038

Note: Total PPI-R, SCI, FD and C represent the *t*-scores for each variable. Positive SL = Positive Super-Learning Ratings; Negative SL = Negative Super-Learning Ratings.

*** $p < .001$; ** $p < .01$; * $p < .05$.

As significant correlations were found between psychopathy traits and inhibitory, but not excitatory, learning the potential difference in relationship between psychopathic traits and excitatory and inhibitory learning was explored by comparing the correlation between psychopathic traits and A Shift Ratings and B Shift Ratings. This was done by using Fisher's *r*-to-*z* transformation and then using a z-test to compare the two correlations (I. A. Lee & Preacher, 2013). The difference in correlations was not significant for the overall psychopathy traits or the factors of SCI and FD. This suggests that there is less evidence that excitatory shift learning is related to psychopathy, even though this comparison is not significant.

As can be seen in Figure 11, higher FD traits were predictive of increased shift learning to the inhibitory cue [A; β = .91, 95% CI (.21, 1.61), $t(46) = 2.62$, $p = .012$, η_p^2 = .13, 95% CI (.01, .31)]. In addition, the total effect of FD on positive super-learning (the sum of the direct effects and indirect effects of FD) was also significant [β = -.36, 95% CI (-.98, -.05)]. This sum of the direct and indirect effects indicates as FD traits increase, the extent of positive super-learning decreases, which is consistent with the direct relationship between FD and shift ratings to A on which the super-learning is dependent. No other direct or indirect effects were significant.

Also shown in Figure 11, higher FD traits predicted greater shift errors to the excitatory cue [B; β = .07, 95% CI (.02, .12), $t(46) = 2.60$, $p = .013$, η_p^2 = .13, 95% CI (.01, .31)]. Shift ratings to the excitatory cue (B) predicted reduced negative super-learning [β = -.44, 95% CI (-.75, -.14), $t(46) = -2.94$, $p = .005$, η_p^2 = .16, 95% CI (.02, .34)]. Male participants also showed smaller shift ratings [β = -21.93, 95% CI (-40.73, -3.13), $t(46) = -2.35$, $p = .023$, η_p^2 = .11, 95% CI (.01, .28)], but greater shift errors [β = 1.06, 95% CI (.05, 2.06), $t(46) = 2.12$, $p = .040$, η_p^2 = .09, 95% CI (.00, .26)] to the excitatory cue (B). However, the 95% confidence intervals about the effect size of the

difference in errors crossed zero, and so this suggests there may in fact be no effect of sex on shift errors to B. None of the direct or indirect effects were significant, all *β*s < .34, all 95% CIs (< -.15, > .11).

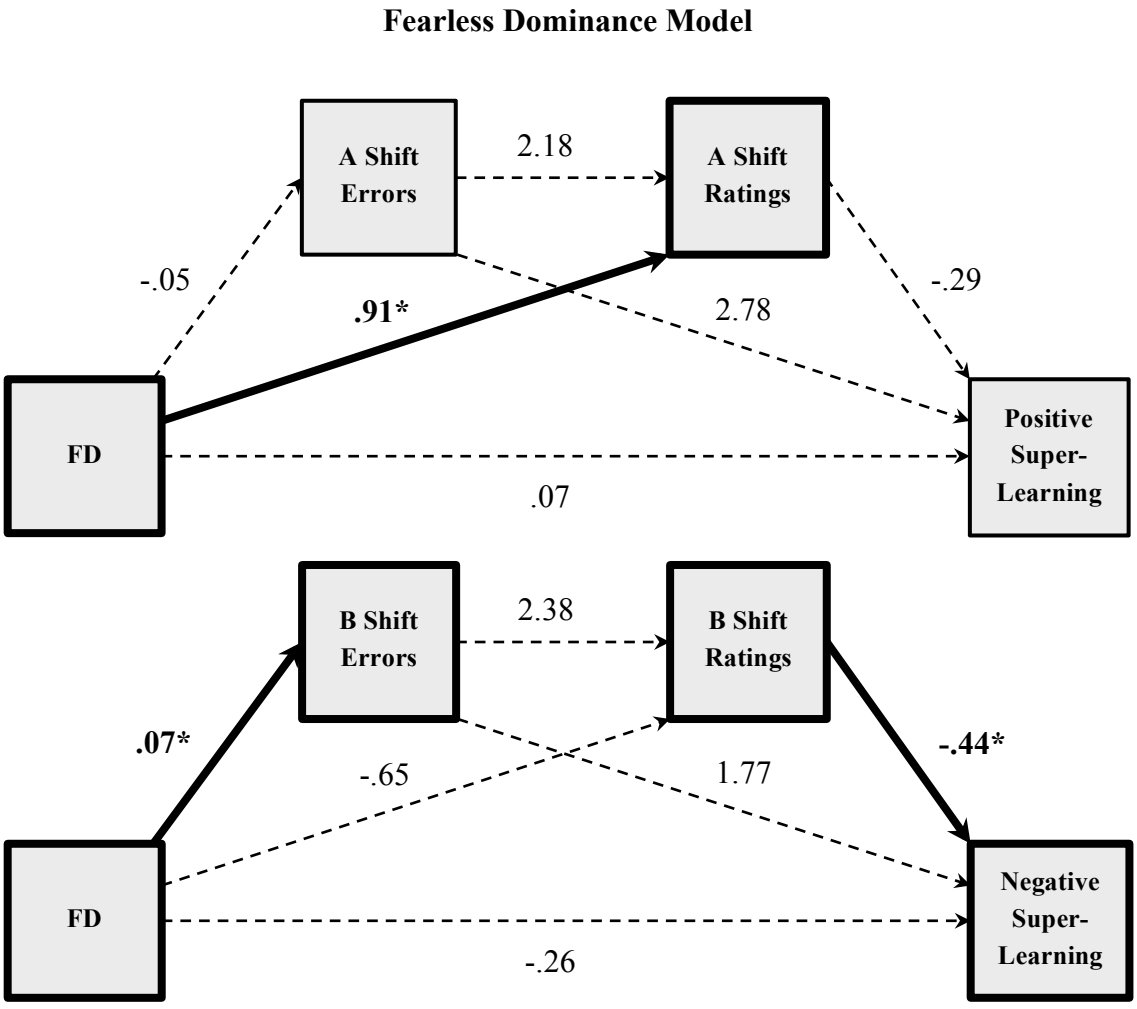


Figure 11. The numbers shown represent *β* values, signalling the strength of a relationship between two variables. FD positively predicted shift ratings to the inhibitory cue (A), but not to the excitatory cue (B). In addition, the total effect of FD on positive super-learning was significant. Shift ratings to the excitatory cue (B) negatively predicted negative super-learning, and FD positively predicted shift errors to the excitatory cue (B).
* $p < .05$

The total effect of SCI on positive super-learning was significant [β = -.26, 95% CI (-.88, -.00)], indicating that as SCI traits increase the extent of positive super-learning decreases. In addition, shift ratings to the inhibitor (A) mediated the relationship between SCI and positive super-learning [β = -.21, 95% CI (-.99, -.00)]. The analysis also showed that female participants made greater shift ratings to the inhibitor (A) than did the male participants [β = 15.08, 95% CI (1.45, 28.71), $t(46) = 2.23$, $p = .031$, $\eta_p^2 = .10$, 95% CI (.00, .27)]. However, consistent with the analysis of variance, as the 95% confidence intervals about the effect size crossed zero, suggesting that there may in fact have been no difference in shift ratings to A between the male and female participants. All other direct and indirect effects were not significant.

Finally, SCI did not directly or indirectly predict negative super-learning [all βs < .32, all 95% CIs (<-.07, >.05)], or either of the mediating variables. As with the FD mediation analysis, male participants showed significantly greater shift learning to the excitatory cue [B; β = -24.87, 95% CI (-43.38, -6.35), $t(46) = -2.71$, $p = .010$, $\eta_p^2 = .14$, 95% CI (.01, .32)], and increased shift learning to the excitatory cue (B) predicted reduced negative super-learning [β = -.43, 95% CI (-.73, -.14), $t(46) = -2.94$, $p = .005$, $\eta_p^2 = .16$, 95% CI (.02, .34)].

3.4.3.3. *Post Hoc* Analysis of the Inhibitory Cue

The mediation analysis results showed that both FD and SCI predicted increased shift learning to the inhibitory cue (A). It is possible that this is due to a failure to acquire strong inhibitory associations in the acquisition phase, which subsequently makes shifting this association easier for high psychopathy trait individuals than it is for their low psychopathy trait counterparts who do acquire inhibition. As such, *post hoc* analysis was conducted to test whether the ratings of the trained inhibitor (A; AX-, X+) were

significantly lower than those of the cue paired with the absence of the outcome (V; BV+, V-), and whether this difference was inversely related to FD and SCI.

As discussed in Chapter 1 (1.7.), previous research has shown that these two procedures can train different strengths of inhibition, and that exposure to V- might result in a reduced likelihood of responding, but that this does not necessarily involve inhibitory processes (McLaren & Verbruggen, 2015). Therefore, it was predicted that the cue that received inhibitory training (A; AX-, X+) will have a significantly lower rating than the cue paired with the absence of the outcome, but which did not receive inhibitory training (V; BV+, V-). This comparison of A and V was done to assess whether individuals with high FD and SCI had acquired associative inhibition in the acquisition phase, as this was a possible explanation for the enhanced shift learning they had shown.

This *post hoc* analysis was done by conducting a 2 (cue: A vs. V) x 2 (flavour counterbalancing) x 2 (sex) mixed model ANOVA, which confirmed that ratings of A (M = -48.89, SD = 20.90) were significantly lower than ratings of V (M = -27.09, SD = 24.73), $F(1, 43) = 29.80$, $p < .001$, $\eta_p^2 = .41$, 95% CI (.18, .57). A simultaneous multiple linear regression analysis was conducted to determine whether this difference was significantly predicted by either FD or SCI traits. As with all the other analysis, flavour counterbalancing and sex were included in the model. Though overall the regression model was not significant, $R^2_{adjusted} = .371$, $F(4,46) = 1.67$, $p = .174$, higher FD traits, but not higher SCI traits, significantly predicted a reduced difference between ratings of A and V, $\beta = -.344$, 95% CI (-1.78, -.10), $p = .029$. All other βs < .21, all other ps > .17.

3.4.4. Discussion

This experiment showed that: psychopathic traits are predictive of differences in learning about inhibitors but not excitors; and, that psychopathic traits are predictive of more accurate responding to an excitatory cue but similar responding to an inhibitory cue. Comparison of ratings across the acquisition and shift phases offered a direct assessment of learning, whilst the ability of the test cue to facilitate enhanced learning about another cue (e.g., super-learning; Rescorla, 1971b; B. A. Williams & McDevitt, 2002) provided an indirect test of the two test cues' associative strength. In both these direct and indirect tests FD and SCI were predictive of differences in learning about the inhibitory cue: high levels of FD and SCI predicted extensive retraining of the previously established inhibitor, and consequently predicted reduced positive super-learning, which was based on the inhibitor.

It appears that the more extensive shift that was seen in individuals with high psychopathic traits meant that the inhibitor had a weak inhibitory value in the shift phase, and so was less able to facilitate the enhanced excitatory learning. In contrast, the weaker shift that was seen in individuals with low psychopathic traits means that the inhibitor had a strong inhibitory value in the shift phase, and so was more able to facilitate enhanced excitatory learning. As such, the relationship between super-learning and psychopathic traits is consistent with the direct assessment of the inhibitor. Psychopathic traits were not predictive of differences in either shift learning or super-learning to the previously established excitor. However, FD traits did predict increased difficulty in learning to reduce responding to a cue that previously required a response, as measured by accuracy of behaviour.

There are at least two possible interpretations of this data: first, that psychopathic traits are predictive of altered cue associability, which precludes the interference of

previous learning or learning about other cues; and second, that psychopathic traits are positively associated with altered stimulus-outcome contingency sensitivity. The former explanation relates to the principles of learning believed to govern the acquisition and shift of associations (e.g., Mackintosh, 1975). As previously discussed, some accounts suggest that learning will be strongest when a cue is predictive of an outcome (Mackintosh, 1975). According to Mackintosh's theory, a degree of cue competition governs learning, and a cue's associability is determined by its relative validity as a predictor of the outcome. This process of determining a cue's relative validity is dependent on integrating information about the predictiveness of all the other cues associated with the outcome. More predictive cues have a higher relative validity, and so a higher associability, meaning that they are more easily learnt about.

Mackintosh's (1975) theory predicts that there will be a decrease in the two test cues' αs in the shift phase because they are inconsistently paired with their respective outcomes. However, at the start of the shift phase, α_A, for instance, will still be high given its predictive strength in the acquisition phase, and so this will initially facilitate learning about the altered contingencies (Mackintosh & Holgate, 1968). As the trials progress α_A will decrease, and so learning will diminish. Therefore, Mackintosh's model predicts that participants should acquire inhibitory strength in the acquisition phase, and that there will be moderate shift learning whilst the α remains high. This is consistent with the performance of the low psychopathy trait individuals, and the performance seen overall for the excitatory cue. If psychopathy is characterised by a problem in integrating information, this process of determining a cue's relative validity will be impaired. As such, in participants with high psychopathy traits the value of α will not adjust to reflect the cues' relative validity with the same degree of sensitivity that is seen in non-psychopathic individuals, which will result in more extensive shift learning.

The difference in excitatory and inhibitory learning can also be accounted for by a failure to integrate information, and the effect that this has on determining a cue's associability. As previously discussed: associative inhibition is dependent on the ability to integrate learning about two different stimuli: e.g., AX- and X+ (McLaren & Verbruggen, 2015; Rescorla, 1969); whilst associative excitation is dependent on learning about one stimulus only, e.g., BV+. Therefore, when determining an inhibitory cue's relative validity there are more cues that need to be integrated into the process than there are when the relative validity of an excitatory cue is in question. If, as has been suggested, psychopathy is characterised by an attentional and perceptual bottleneck (Hamilton et al., 2015), then the process of determining a cue's relative validity and so adjusting its associability may be impaired. In addition, integration of information is required in order to develop associative inhibition. As fearless dominance was negatively predictive of a difference in ratings to the associative inhibitor (A) and a cue simply paired with the absence of the outcome (V), this suggests that psychopathic individuals may also struggle to form the initial inhibitory association.

However, there is an alternative explanation for this data: that psychopathic traits are related to an increased sensitivity to contingencies. In the shift phase, the contingency between test cue A and the outcome was changed and though this was reflected in the significant main effect of phase, sensitivity to this change was predicted by psychopathic traits. By suggesting that psychopathic traits are predictive of increased contingency sensitivity, it is also possible to account for the imbalance between excitatory and inhibitory learning in the present study.

Experiment 3 replicated some of the findings of Experiment 2 by appearing to show the novel effect of so-called 'negative super-learning', as well as positive super-learning. However, it is possible that these results are not simply the consequence of

super-learning but rather also reflective of overshadowing or blocking (Kamin, 1969) of the control cues. Overshadowing is described when a cue's acquisition of associative strength is weakened due to its co-occurrence with a cue that is more predictive, whilst blocking occurs when a cue that is predictive of the outcome prevents another cue with which it is reinforced from acquiring associative strength as the original (blocking) cue is already predictive of the outcome. Therefore, like the super learning effects, the strength of the association between a cue and an outcome is influenced by the strength of the other cues with which it is paired. For example, in this experiment cue D was paired with the outcome both as an element (D+), and when in compound with the super-learning comparison cue S (DS+). Therefore, it is possible that D was more predictive and so overshadowed or interfered with the excitatory association developed by S. This draws in to question the appropriateness of the control cues in both the positive and negative super-learning tests as the value of the control cues may be compromised.

Either of these processes of overshadowing or blocking (Kamin, 1969) would have increased the likelihood of finding a significant difference between the positively super-learnt cue (Y) and the comparison cue (S) by reducing the associative strength acquired by S. The same is true of the negatively super-learnt cue and its comparison cue. As such, these results cannot be argued to unequivocally show super-learning and future studies with control cues more similar to those seen in Experiments 1 and 2 are needed (e.g., DS+ or DS+, D-). This brings into doubt whether the evidence is purely based on super-learning. However, it is still consistent with showing that learning about a novel cue reinforced in compound with an established predictor of the same outcome (e.g., D+ DS+; C-, CZ-) is different to learning about a novel cue that is reinforced in compound with an established predictor of the opposite outcome (e.g., A-, AY+; B+, BW-). A

systematic replication of these effects using a modified design was carried out in Experiment 4.

Finally, it is noted that female participants showed greater shift learning to the excitatory cue, but not to the inhibitory cue, suggesting that females are better at learning that a cue that previously signalled an outcome no longer does. That is to say that female participants learnt the devaluation of a cue more easily than their male counterparts.

To conclude, psychopathic traits predict differences in shift learning and subsequent super-learning about inhibitory but not excitatory cues. This may be attributed to either a failure to integrate information necessary for adjusting a cue's associability or an increased sensitivity to contingencies. Finally, the results of Experiment 3 appear to provide a replication of findings from Experiment 2 by demonstrating the novel negative super-learning effect, though a replication of these effects using a more stringent control will be conducted in Experiment 4.

3.5. Experiment 4

3.5.1. Introduction

Experiment 3 showed that psychopathic traits are related to fragile learning about inhibitory cues, resulting in greater shift when the stimulus-outcome contingencies were changed, and consequently reduced second-order effects based on the strength of this inhibitor. These differences were not seen for the excitatory cue and its second-order effects. However, it is not clear whether this reflects a difference in the way in which excitors and inhibitors are learnt about and how this might affect their associability (Mackintosh, 1975), or whether it reflects a difference in contingency sensitivity.

So, in this experiment the learning design was adjusted to ensure that the change in associative strength to the excitatory and inhibitory cues was equal. Whilst X and V, the cues with which test cues A and B respectively were trained in the acquisition phase, were presented as elements in Experiment 3 (AX-, X+; BV+, V-) in the new design they were paired in compound with neutral stimuli (AX-, DX+, D+/-; BV+, CV-, C+/-). Training in the shift phase remained the same balanced design used in Experiment 3.

As was also discussed in Experiment 3 (3.4.4.), it is possible that the evidence of super-learning was influenced by either an overshadowing or a blocking effect (Kamin, 1969), thereby reducing the associative strength acquired by the comparisons cues. This would have increased the likelihood of finding significant super-learning effects. So, Experiment 4 sought to replicate and extend the results of Experiment 3 by adjusting the experimental design to reduce potential overshadowing or blocking, whilst keeping the similarity of training between the super-learnt and comparison cues in the shift phase. To this end, cues C and D, which were trained in compound with the super-learning comparison cues in the shift phase, were inconsistently paired with the outcome in acquisition phase (C+/- and D+/-) so as to reduce their predictiveness, and so limit the

associative strength they acquired in the shift phase. This was intended to reduce the likelihood of any potential overshadowing or blocking effects occurring.

3.5.2. Method

3.5.2.1. Participants and Design

Forty-seven (21 male, 26 female; age $M = 21.72$ years, $SD = 3.05$) individuals participated using the same recruitment criteria, design and ethical approval as in Experiments 2 and 3. As in Experiment 3, power analysis conducted on the basis of the negative super-learning in Experiment 2 indicated that 47 participants would be needed to have 80% power for detecting an effect when employing the traditional .05 criterion of statistical significance.

3.5.2.2. Materials

Psychopathic Personality Inventory (Revised; PPI-R; Lilienfeld & Widows, 2005). In the present sample, the internal consistency of the PPI-R was good ($\alpha = .91$). The means and standard deviations for the participants are shown in Table 10.

Table 10

Experiment 4 Participant Psychopathy *T*-Scores

	M	*SD*
Total PPI-R	49.36	8.98
Self-Centred Impulsivity	52.66	9.57
Fearless Dominance	45.85	8.89
Coldheartedness	48.62	10.78

Note: N = 47. The mean *t*-score is 50.

Shift Go/NoGo Task. As shown in Table 11, the design was changed to reduce potential overshadowing or blocking (Kamin, 1969) of the super-learning comparison cues.

Table 11

Experiment 4 Design

Acquisition Phase	**Shift Phase**
AX-	AY+, A-
BV+	BW-, B+
CV-, C+/-	CZ-, C-
DX+, D+/-	DS+, D+

Note: Overall outcome direction is shown as excitatory (+) or inhibitory (-). A and B are the shift test cues, and Y and W the super-learnt test cues. All the other cues are included to counterbalance the training of A, B, Y and W.

Cues C and D, which were trained in compound with the super-learning comparison cues in the shift phase, were inconsistently paired with the outcome in the acquisition phase (C+/- and D+/-) so as to reduce their predictive validity, and their potential to overshadow or block the super-learning comparison cues in the shift phase. In addition C was paired with V (CV-) so as to reduce the moderate super-learning of B when it is paired with V (BV+, CV-, C+/-). So as to balance exposure to elements and compounds, D was paired with X (DX+) the excitatory cue used to train the inhibitor A (AX-, DX+, D+/-).

3.5.3. Results

As shown in Figure 13, participants were able to learn and shift their learning about excitatory and inhibitory cues. SCI predicted increased shift learning about the inhibitor (A), but not the excitor (B). As shown in Figure 14, participants showed significant positive and negative super-learning, but this was not related to psychopathic traits. Consistent with Experiment 3, female participants showed greater shift learning to the excitatory cue than the male participants did.

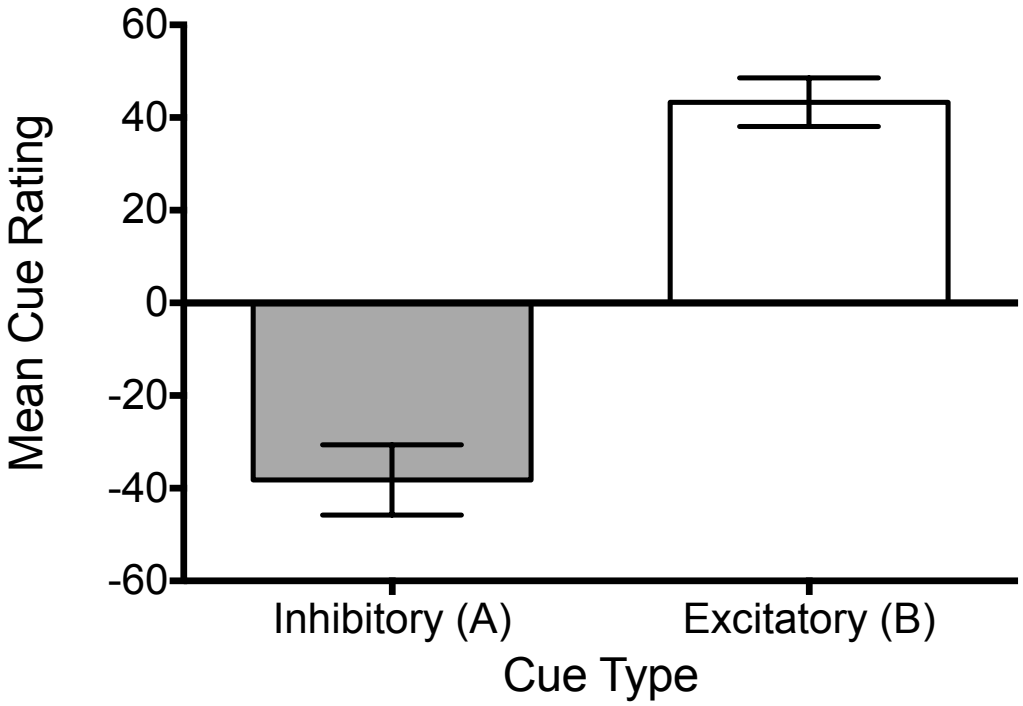


Figure 12. Participants rated the inhibitory cue (A) as significantly more negative than the excitatory cue (B) in the acquisition phase. Error bars represent 95% confidence intervals.

3.5.3.1. Overall Learning Effects

Initial Cue Learning (Acquisition Phase). As shown in Figure 12 participants

rated the inhibitory cue (A; *M* = -38.20, *SD* = 25.74) as significantly more negative than the excitatory cue (B; *M* = 43.32, *SD* = 17.90), $F(1, 43) = 270.33$, $p < .001$, $\eta_p^2 = .86$, 95% CI (.77, .90). No interactions were significant, all *F*s < 2.39, all *p*s > .12.

Shift Errors. Overall, participants did not make a significantly different number of errors in the acquisition phase (*M* = .39, *SD* = .58) and the shift phase (*M* = .57, *SD* = .75) for test cue A, all *F*s < 2.6, all *p*s > .11, and for test cue B (acquisition phase: *M* = .57, *SD* = 1.14; shift phase *M* = .51, *SD* = .75), all *F*s < 2.0, all *p*s > .17.

Shift Ratings. As shown in Figure 13, participants shifted their learning to the previously predictive inhibitor (A; acquisition: *M* = -38.21, *SD* = 27.13; shift: *M* = -25.75, *SD* = 28.37), $F(1,43) = 10.33$, $p = .002$, $\eta_p^2 = .19$, 95% CI (.03, .38). No interactions were significant, all *F*s < 2.18, all *p*s > .14. Likewise, participants shifted their learning to the previously predictive excitor (B; acquisition: *M* = 43.32, *SD* = 17.90; shift: *M* = 27.97, *SD* = 18.99), $F(1,43) = 43.54$, $p < .001$, $\eta_p^2 = .50$, 95% CI (.28, .64). Female participants showed greater shift to the excitor than the male participants did, $F(1,43) = 9.33$, $p = .004$, $\eta_p^2 = .18$, 95% CI (.02, .37). *Post hoc t*-tests carried out to compare excitatory shift learning in males and females were significant for both the females, $t(25) = 7.10$, $p < .001$, $\eta_p^2 = .67$, 95% CI (.41, .78), and the males, $t(20) = 2.65$, $p = .015$, $\eta_p^2 = .26$, 95% CI (.01, .51), though the effect size was larger in the female participants. No other interactions were significant, all *F*s < 1.76, all *p*s > .19.

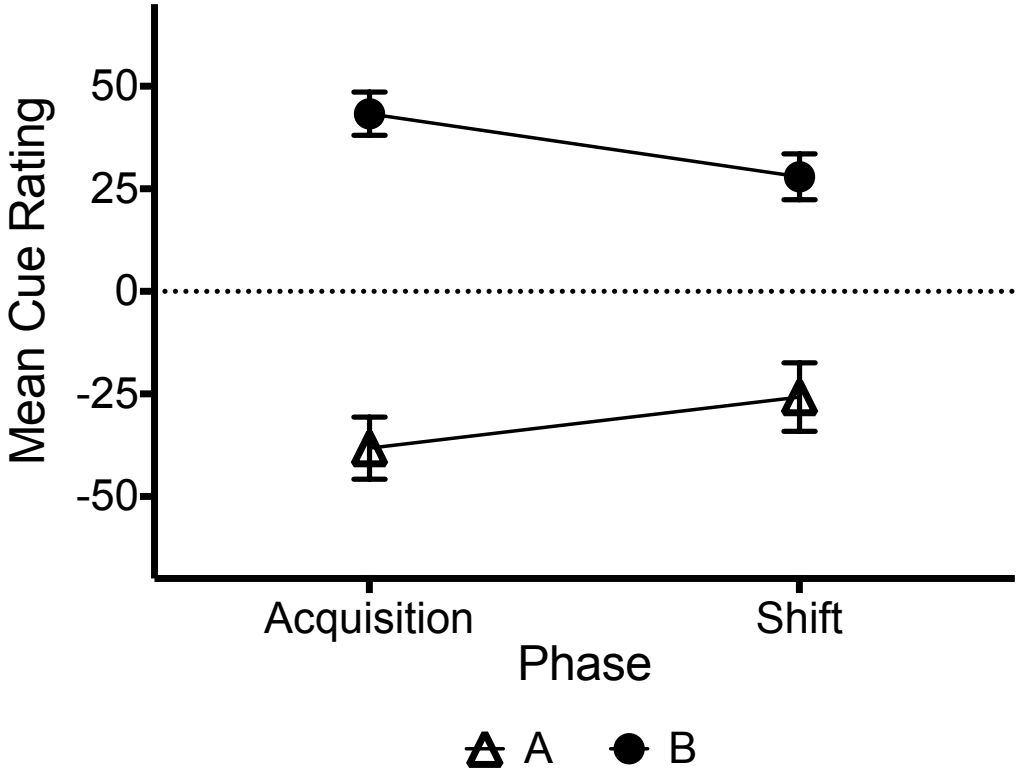


Figure 13. Participants showed significant shift to the inhibitory (A) and excitatory (B) test cues. Error bars represent 95% confidence intervals.

Super-Learning. Overall participants showed enhanced learning to Y (M = 45.28, SD = 29.06) as compared to S (M = 26.04, SD = 30.18), so-called positive super-learning, $F(1,43) = 11.64$, $p = .001$, $\eta_p^2 = .21$, 95% CI (.04, .40). No interactions were significant, all Fs < 2.71, all ps > .10. Likewise, participants also showed enhanced learning to W (M = -46.27, SD = 23.39) as compared to Z (M = -22.16, SD = 24.47), so-called negative super-learning, $F(1,43) = 34.11$, $p < .001$, $\eta_p^2 = .44$, 95% CI (.21, .60). No interactions were significant, all Fs < 2.10, all ps > .15. These main effects are shown in Figure 14.

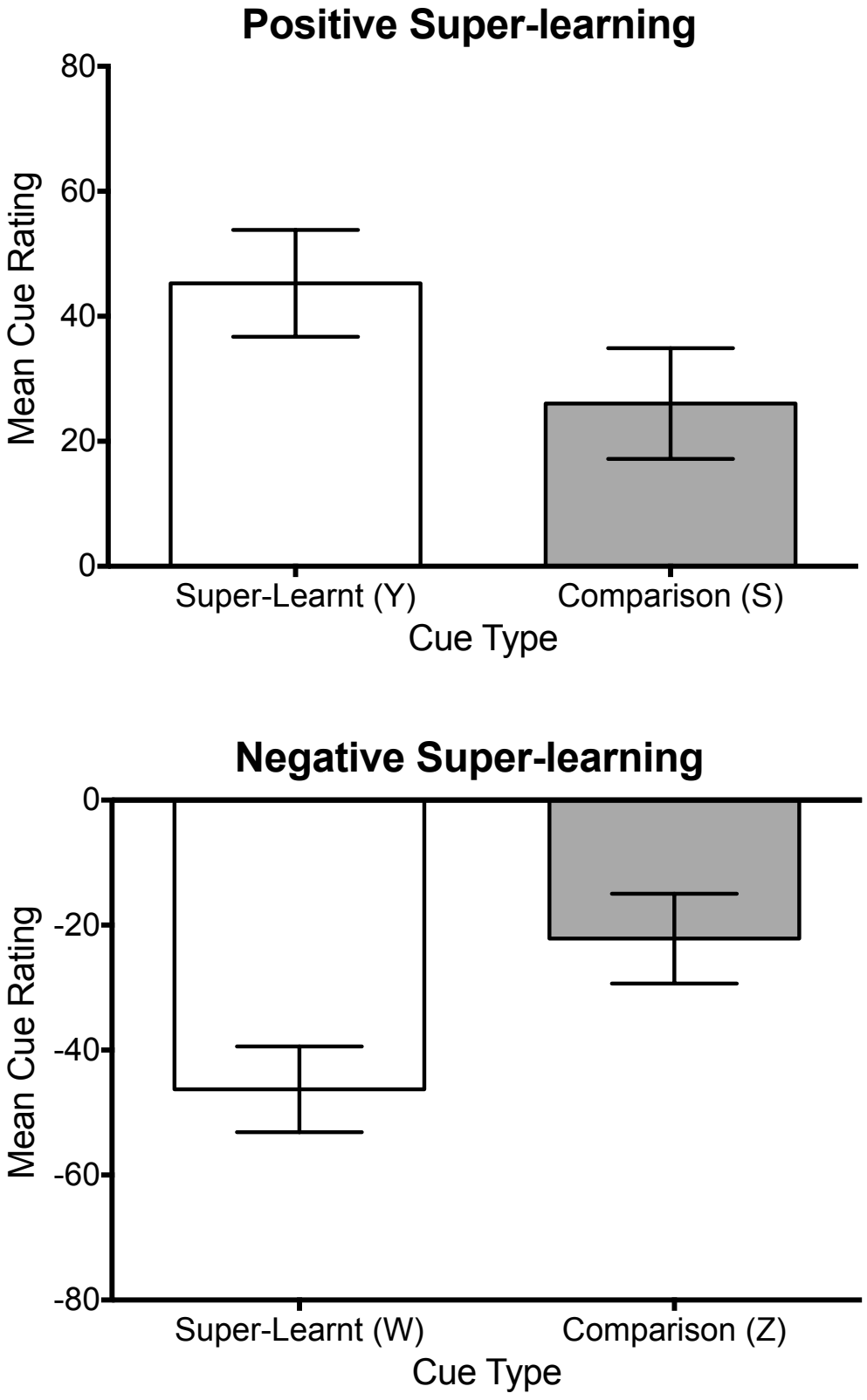


Figure 14. Participants showed significant super-learning to Y and W as compared with S and Z respectively. Error bars represent 95% confidence intervals.

3.5.3.2. Relationship Between Psychopathy and Dependent Variables

The zero-order correlations between overall psychopathy scores, the factors of SCI, FD and C, and the dependent variables are shown in Table 12. Once again, each of the three PPI-R factors was strongly related to the overall PPI-R score. SCI was significantly correlated with the shift in ratings to the inhibitory test cue (A; r = -.32, p = .028), and the difference in errors to the inhibitor (A; r = -.30, p = .041). All other dependent variables were not significantly related to psychopathy.

Table 12

Experiment 4 Zero-Order Correlations

	Total PPI-R	**SCI**	**FD**	**C**
Total PPI-R	**-**			
SCI	**.821*** **	-		
FD	**.529*** **	.005	-	
C	**.448** **	**.345***	-.067	-
A Shift Errors	-.203	**-.300***	.019	.085
B Shift Errors	.053	.032	-.149	.052
A Shift Ratings	.251	**.321***	.009	.023
B Shift Ratings	.104	.052	.175	-.118
Positive SL	-.022	-.123	.128	.031
Negative SL	-.243	-.213	-.076	-.204

Note: Total PPI-R, SCI, FD and C represent the *t*-scores for each variable. Positive SL = Positive Super-Learning Ratings; Negative SL = Negative Super-Learning Ratings.

*** p < .001; ** p < .01; * p < .05.

As in Experiment 3, the difference in correlations between psychopathic traits and excitatory and inhibitory learning were tested (I. A. Lee & Preacher, 2013). Consistent with the results of Experiment 3, the difference between excitatory and inhibitory correlations with the factor of SCI were not significant.

FD traits did not directly or indirectly predict either positive super-learning [all βs < .60, all 95% CIs (<-.07, >.02)] or negative super-learning [all βs < .27, all 95% CIs (<-.02, >.05)]. Female participants showed greater shift ratings to the excitatory cue (B) than did their male counterparts, [β = 14.22, 95% CI (5.19, 23.25), $t(46) = 3.18$, $p = .003$, η_p^2 = .18, 95% CI (.02, .36)]. SCI traits did not directly or indirectly predict either positive super-learning [all βs < .38, all 95% CIs (<-.03, >.26)] or negative super-learning [all βs < .69, all 95% CIs (<-.03, >.01)]. However, SCI traits did significantly predict increased shift ratings to the inhibitory cue [A; β = .99, 95% CI (.16, 1.82), $t(46) = 2.41$, $p = .020$, η_p^2 = .11, 95% CI (.01, .29)]. In addition female participants showed significantly greater shift ratings to the excitatory cue [B; β = 13.94, 95% CI (4.70, 23.18), $t(46) = 3.04$, p = .004, η_p^2 = .17, 95% CI (.02, 35)].

3.5.3.3. *Post Hoc* Analysis of the Inhibitory Cue

Given the significant relationship between shift learning the inhibitory cue (A) and SCI traits, *post hoc* comparison of participants' ratings of A and V in the acquisition phase was carried out as in Experiment 3. A 2 (cue: A vs. V) x 2 (flavour counterbalancing) x 2 (sex) mixed model ANOVA confirmed that ratings of A (M = -38.21, SD = 25.74) were significantly lower than the ratings of V (M = -2.68, SD = 25.07), $F(1, 43) = 57.66$, $p < .001$, η_p^2 = .57, 95% CI (.36, .69). A simultaneous multiple linear regression analysis including FD, SCI, sex and flavour counterbalancing was not significant, all βs < .13, all ps > .43.

3.5.4. Discussion

Experiment 4 broadly replicated the findings of Experiment 3, and helps to clarify some of the questions raised in Experiment 3. As shown previously, psychopathic traits were not related to behavioural response errors, but they were related to a greater shift in associative strength to the previously trained inhibitor. This study used a new design to tightly control for any biases in the magnitude of change in associative strength to the cues. Despite this, psychopathic traits were still not related to a difference in shift learning to the excitor, but remained predictive of differences in shift learning to the inhibitor. This supports the specificity of the difference in relation to psychopathic traits that was shown in Experiment 3, and suggests that it was not due to the magnitude of change in associative strength, as could have previously been argued. Instead, this difference is related specifically to inhibitory cues, a finding consistent with previous research showing impaired inhibition in psychopathic samples (He et al., 2011).

This inhibitory specificity of the psychopathy effect is supportive of an associability account rather than a contingency sensitivity account. Whilst a contingency sensitivity explanation would have predicted symmetrical psychopathy effects about the inhibitor and the excitor, the associability account predicted psychopathy differences in the inhibitor only. As outlined in the discussion section of Experiment 3 this is because in order to determine a cue's relative validity, and so its associability (something relevant to both excitatory and inhibitory cues) the integration of information is required (Mackintosh, 1975). The demands of this process will be greater for inhibitory cues which themselves require the integration of information in their formation. Limited processing capacity (Hamilton et al., 2015) means that psychopathic individuals may be unable to integrate all the necessary information, resulting in enhanced shift learning to the inhibitor.

The *post hoc* analysis of the inhibitor suggested that psychopathic traits were not predictive of difficulty in acquiring the initial inhibitory association, as was suggested by the results of Experiment 3. Therefore, in order to account for the inhibitory specificity, an amendment to the framework of Mackintosh's (1975) model is required. Mackintosh suggested that the relative validity of a cue is determined 'locally', meaning that all trials present in a series would not be considered. However, to account for the inhibitory specificity, I propose that the number of trials taken in to consideration is larger than previously suggested, such that the inhibitory training of A in the acquisition phase is still affecting the determination of the cue's relative validity in the shift phase. Through this small amendment to the model, it is possible to explain the inhibitory specificity, without the presence of altered inhibitory learning in the acquisition phase.

Unlike in Experiment 3, this difference in inhibitory shift learning was only predicted by SCI traits, and not by FD traits. However, this particular sample scored relatively low on levels of FD and relatively high on levels of SCI. Therefore, it is likely that participants' FD scores did not have a sufficient range to facilitate true testing of whether they predicted differences in the participants' learning. Another difference compared to Experiment 3 is that neither SCI nor FD directly or indirectly predicted super-learning. The absence of this effect in relation to FD can be accounted for as above, however this does not account for the absence of an effect between SCI and super-learning. This is particularly surprising given that SCI predicts shift ratings of A, the cue on which super-learning is dependent.

It is possible that changes to the design are responsible for this difference. It was discussed in Experiment 3 that the observed super-learning effects might be attributable to either overshadowing or blocking (Kamin, 1969) of the comparison cues (i.e., S and Z) by the cues that they were trained with (i.e., D and C respectively). To minimise this

possibility, whilst retaining the balanced shift phase design, cues C and D that were potentially doing the overshadowing or blocking were pre-exposed in the acquisition phase to reduce their predictiveness. As such, the likelihood of either blocking or overshadowing occurring was reduced, as the associability of the pre-exposed cues was low, suggesting that the comparison cues would be able to acquire excitatory and inhibitory associative strength. That the psychopathy effects appear to relate to this design change suggests that participants with low psychopathic traits were particularly susceptible to overshadowing or blocking effects in Experiment 3, which is why their super-learning appeared to be stronger; whilst the high psychopathy trait individuals' learning was less affected by the potential blocking or overshadowing occurring in Experiment 3.

This suggestion that psychopathic traits were related to reduced interference by other cues is consistent with the conceptualisation of the observed learning differences resulting from a failure to integrate information: overshadowing and blocking occur because another cue is either a more salient, or a better predictor of the outcome; whilst super-learning occurs because a cue that is predictive of one outcome is paired with a novel cue and trained as predictive of the opposite outcome. Therefore, all three processes depend on the integration of information across cues.

If psychopathic individuals have a narrow processing bottleneck and a difficulty integrating information, then they will be less likely to show enhanced learning (e.g., super-learning), or reduced learning (e.g., overshadowing or blocking) when the processes depend on combining information across cues. This may explain why psychopathic traits predicted reduced positive super-learning in Experiment 3, where the comparison cue was potentially overshadowed or blocked, as they were less susceptible to such interference, and so rated the super-learnt and comparison cues more similarly.

However, in the present experiment the likelihood of such overshadowing or blocking occurring was reduced through the changes to the design, meaning that the difference in the super-learning effect was not dependent on overshadowing or blocking of the comparison cue, and so all participants showed super-learning. Despite these changes in the design intended to reduce the likelihood of blocking or overshadowing occurring, it is nonetheless possible that the observed differences in cue ratings believed to show super-learning in fact are driven by the suppression of the comparison cues, as opposed to enhanced learning about the intended super-learnt cues. Given the design used here, it is not possible to exclude this possibility, and so these results must be interpreted with caution until further studies using an amended design, such as that used in Experiment 2, can be used to clarify these issues.

Of interest is the finding that the shift in learning about the excitatory cue, whilst not related to psychopathic traits, was significantly different for male and female participants. As in Experiment 3, female participants showed a greater shift in their excitatory test cue associations when the cue's relationship with the outcome was weakened. In contrast, male participants showed a weaker ability to learn that a cue that was previously a strong excitatory signal became a weaker excitatory signal.

In summary, the results of Experiment 4 have broadly replicated those of Experiment 3, and have clarified that the inhibitor-specific difference that is associated with psychopathic traits is not due to a disparity in the magnitude of change in associative strength. Instead, the present findings support the case that psychopathic traits are associated with specific difficulty in integrating information (Hamilton et al., 2015), which is argued to prevent the updating of a cue's associability (Mackintosh, 1975). This is a particular problem for inhibitory learning, which requires more information to be integrated than excitatory learning does. Given the relation of this difference to inhibitory

cues, it should be ascertained whether equivalent differences are seen in shift learning to latent inhibitors: cues that are pre-exposed and not predictive of the outcome (Lubow, 1973; Rodriguez & Hall, 2008). If the differences observed in these studies were due to a bias in the rapid acquisition of excitatory associations, then shift learning about latent inhibitors would also be related to psychopathic traits. However, if the inhibitory differences seen here were due to something else central to inhibitors, but not excitors, then a difference would not be seen in relation to psychopathic traits, a hypothesis that will be tested in Experiment 5.

3.6. Experiment 5

3.6.1. Introduction

In order to clarify the findings of the previous studies, this experiment tested participants' ability to distinguish between predictive and non-predictive stimuli. By looking at whether participants could acquire excitatory and inhibitory associations to cues that were previously non-predictive, the associability account that has been proposed could be tested. As such, two cues (C and D) were trained as non-predictive in the acquisition phase, and then as either inhibitory (C-, CZ-) or excitatory (D+, DS+) in the shift phase. Shift learning to these previously non-predictive cues was hypothesised to be unrelated to psychopathic traits. This is because the differences seen in relation to psychopathic traits are present when extensive integration of information is required. The load in determining the relative validity of non-predictive cues is more akin to that of excitatory cues than inhibitory cues, and so was predicted to be within the processing capacity of high trait psychopathy individuals. In addition, the non-predictive cues became more predictive in the shift phase, and so high cue associability was argued to facilitate accurate shift learning. Therefore, all participants were expected to show excitatory and inhibitory shift to the previously non-predictive cues.

The inclusion of these neutral cues in the acquisition phase also allowed a test of the excitatory learning that takes place, by facilitating comparison of an inhibitory, an excitatory and a neutral cue. The results of Experiments 3 and 4 have suggested that psychopathic traits predict altered inhibitory learning, which appears more susceptible to change (shift). This has been interpreted as a demonstration of weaker, or more vulnerable inhibition. However, these differences could also arise from weaker or more disrupted excitatory learning given that an excitatory cue is needed to facilitate the learning about the inhibitory cue. Therefore, if, for example, X+ did not acquire (much)

excitatory strength, then AX- training would not result in A acquiring (much) inhibitory strength. As such, by including a neutral cue in this design, it was possible to assess excitatory association formation by comparing the neutral cue to the excitatory cue, and to assess whether psychopathic traits were associated with differences in this ability, thereby providing a test of whether differences in the formation of excitatory associations was responsible for the observed inhibitory learning differences.

3.6.2. Method

3.6.2.1. Participants

Thirty-six (18 male, 18 female; age M = 23.39 years, SD = 4.17) individuals participated under the ethical and recruitment conditions outlined in the General Method section of this chapter (3.2.2.). A power analysis on the basis of the positive super-learning in Experiment 4 (μ_1 = 45.28; μ_2 = 26.04; σ = 29.02) indicated that 36 participants were needed to have 80% power for detecting an effect when employing the traditional .05 criterion of statistical significance.

3.6.2.2. Design

The design was the same as in Experiments 2-4, with the addition that participants' ratings to C and D in the acquisition and shift phases were compared to assess learning about previously non-predictive stimuli. Simultaneous multiple linear regression analysis, with sex and flavour counterbalancing condition also included was used to test whether psychopathic traits predicted this shift learning.

3.6.2.3. Materials

Psychopathic Personality Inventory (Revised; PPI-R; Lilienfeld & Widows, 2005). In the present sample, the internal consistency of the PPI-R was good (α = .90). The means and standard deviations for the participants are shown in Table 13.

Table 13

Experiment 5 Participant Psychopathy *T*-Scores

	M	*SD*
Total PPI-R	46.22	8.24
Self-Centred Impulsivity	47.64	8.08
Fearless Dominance	46.03	11.20
Coldheartedness	48.36	12.17

Note: N = 36. The mean *t*-score is 50.

Shift Go/NoGo Task. As shown in Table 14, in this design participants were required to learn when to respond and when to inhibit their responses to twelve different stimuli that had positive (excitatory: B+, D+, X+, AY+, BV+, DS+), neutral (T+/-, U+/-, CT+/-, DU+/-) or negative (inhibitory: A-, C-, V-, AX-, BW-, CZ-) contingencies. As in Experiments 2-4, cues A and B were included in the design to test participants' ability to shift their learning to previously predictive cues. Positive and negative super-learning were once again tested by: training the inhibitory cue A with the neutral cue Y (AY+) to facilitate positive learning as compared to S, a novel cue trained in compound as excitatory (DS+); and, by training the excitatory cue B with the neutral cue W (BW-) to facilitate negative learning as compared to Z, a novel cue trained in compound as inhibitory (CZ-).

Cues C and D facilitated testing of whether participants could learn either excitatory or inhibitory associations about cues that were previously trained as non-predictive. C was trained as non-predictive in the acquisition phase, meaning it was not predictive of the outcome (CT+/-). C was then trained as a predictive inhibitory cue in the shift phase (CZ-, C-). Average ratings of C in the acquisition and shift phases were

compared to determine whether shift learning occurred or not to a previously non-predictive cue subsequently trained at inhibitory.

Table 14

Experiment 5 Design

Acquisition Phase	Shift Phase
AX-, X+	AY+, A-
BV+, V-	BW-, B+
CT+/-, T+/-	CZ-, C-
DU+/-, U+/-	DS+, D+

Note: Overall outcome direction is shown as excitatory (+) or inhibitory (-). A and B were initially predictive then reversed on the basis of weakly predictive information, whilst C and D were initially trained as non-predictive, before receiving predictive excitatory and inhibitory reinforcement in the shift phase. All other cues are included to counterbalance the training of A, B, C and D.

D was also trained as non-predictive in the acquisition phase, meaning it was not predictive of the presence or absence of the outcome (D+/-). D was then trained as a predictive excitatory cue in the shift phase (DS+, D+). Average ratings of D at the end of the acquisition and shift phase were compared to determine whether shift learning occurred or not. C and D provided a control for the findings of Experiments 3 and 4, by

assessing whether participants were capable of distinguishing predictive and non-predictive cues.

3.6.3. Results

Consistent with previous experiments, psychopathic traits were not related to participants' errors in responding. Also consistent with Experiments 3 and 4, and as shown in Figure 16, participants showed significant shift learning to both the excitatory and the inhibitory cue. Psychopathic traits significantly predicted the extent of shift seen to the inhibitor. Overall, participants also showed significant shift learning to the previously non-predictive cues, an ability that was not related to psychopathic traits. These findings support and further the results of Experiments 1-4.

3.6.3.1. Overall Learning Effects

Initial Cue Learning (Acquisition Phase). As shown in Figure 15 participants rated the inhibitory cue (A; M = -43.07, SD = 25.89), the neutral cue (C; M = 5.76, SD = 21.60) and the excitatory cue (B; M = 56.26, SD = 24.05) as significantly different from each other, $F(2, 64) = 83.40$, $p < .001$, $\eta_p^2 = .72$, 95% CI (.59, .79). No interactions were significant, all Fs < 1.80, all ps > .17. In order to test whether psychopathic traits predicted differences in excitatory learning, the difference in ratings between the excitatory cue (B) and the neutral cue (C) was calculated by subtracting the rating of the neutral cue from the excitatory cue. This difference score, where a more positive score indicated that the excitatory cue was rated higher than the neutral cue, was then used to test whether previously observed differences in inhibitory learning were due to impaired excitatory association formation. A simultaneous multiple linear regression analysis showed that psychopathy factors were not significantly predictive of the difference in

participants' ratings of an excitatory and a neutral cue, $R^2_{adjusted}$ = -.034, $F(4,35) = .61$, $p = .611$, $\eta_p^2 = .07$, 95% CI (.00, .17) all βs < .19, all ps > .29.

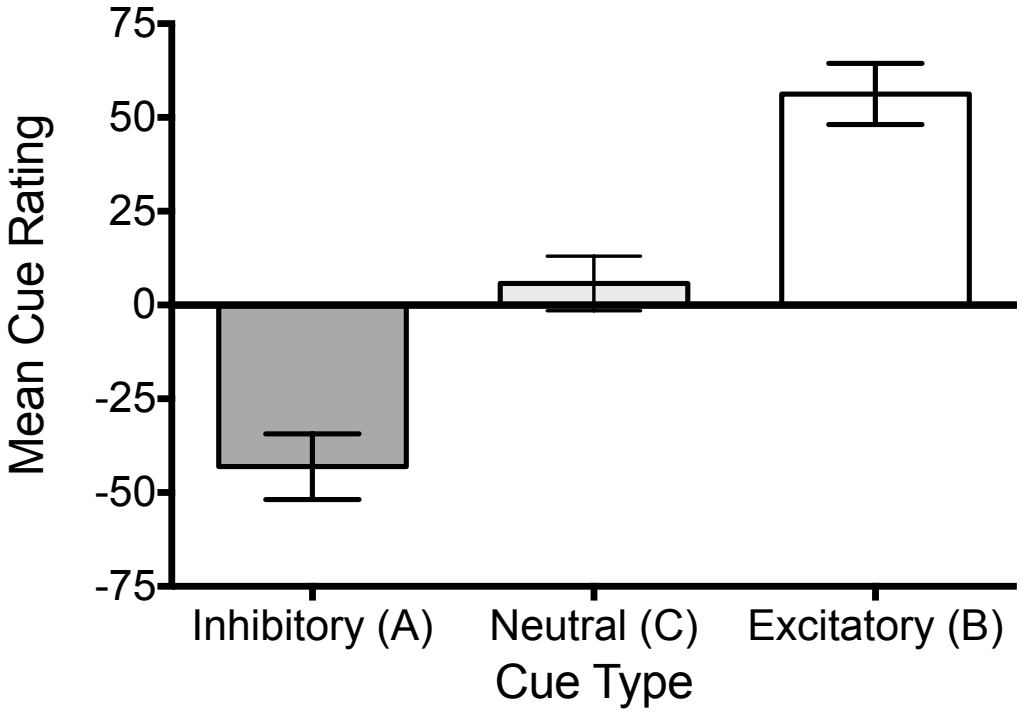


Figure 15. Participants rated the inhibitory cue (A), the neutral cue (C) and the excitatory cue (B) as significantly different from each other. Error bars represent 95% confidence intervals.

Shift Errors (A & B). Overall, participants did not make a significantly different number of errors in the acquisition phase (M = .28, SD = .66) and the shift phase (M = .11, SD = .40) for test cue A, all Fs < 1.26, all ps > .26. Participants also did not make a significantly different number of errors in the acquisition phase (M = .86, SD = 1.87) and the shift phase (M = .86, SD = 1.78) for test cue B, all Fs < .35, all ps > .56.

Shift Ratings (A & B). As shown in Figure 16, overall participants shifted their learning to the test cue A the previously predictive inhibitor (acquisition: M = -43.07, SD

= 25.89; shift: *M* = -18.11, *SD* = 20.32), *F*(1,32) = 28.75, *p* < .001, η_p^2 = .47, 95% CI (.21, .63). No interactions were significant, all *F*s < 4.30, all *p*s > .05. Also shown in Figure 16, participants significantly shifted their learning to test cue B (acquisition: *M* = 56.26, *SD* = 24.05; shift: *M* = 31.52, *SD* = 26.38), *F*(1,32) = 10.26, *p* = .003, η_p^2 = .24, 95% CI (.03, .45). No interactions were significant, all *F*s < .38, all *p*s > .54.

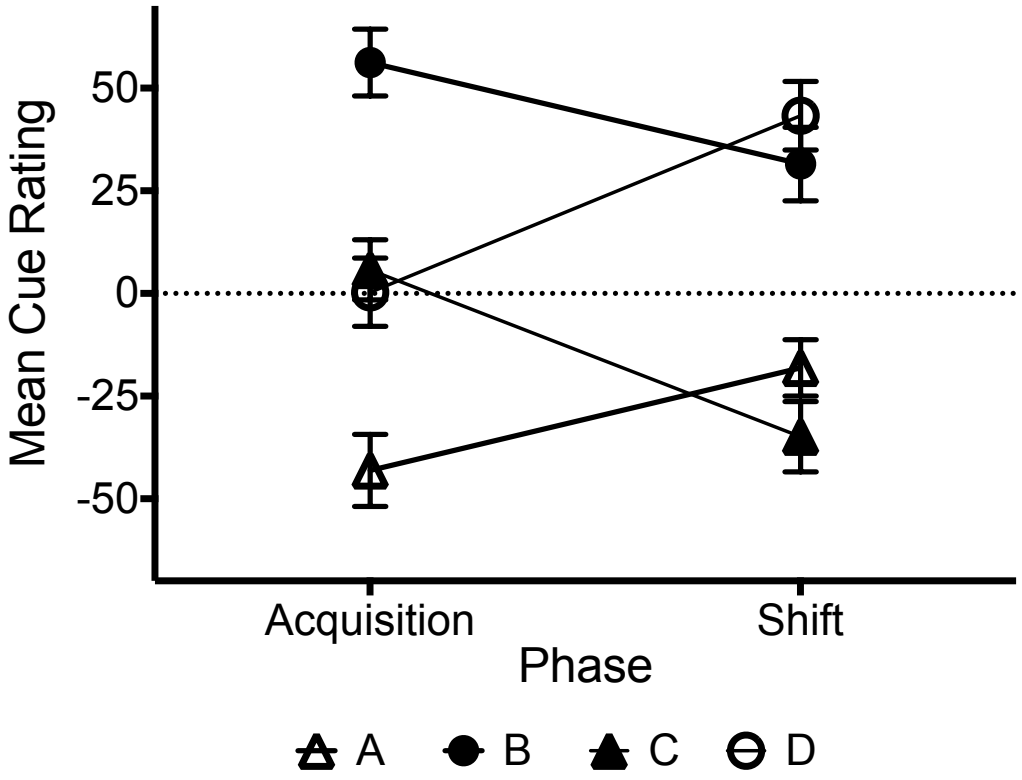


Figure 16. Participants showed significant shift learning the inhibitory cue (A), the excitatory cue (B), and the two non-predictive cues that were subsequently trained as inhibitory (C) and excitatory (D). Error bars represent 95% confidence intervals.

Shift Ratings (C & D). As shown in Figure 16, overall participants shifted their learning to test cue C (acquisition: *M* = 5.76, *SD* = 21.60; shift: *M* = -34.89, *SD* = 25.28),

$F(1,32) = 28.34$, $p < .001$, $\eta_p^2 = .47$, 95% CI (.20, .63). No interactions were significant, all Fs < 1.39, all ps > .17. Also shown in Figure 16, participants shifted their learning to test cue D (acquisition: $M = .31$, $SD = 24.58$; shift: $M = 43.29$, $SD = 24.59$), $F(1,32) = 56.32$, $p < .001$, $\eta_p^2 = .64$, 95% CI (.40, .75). No interactions were significant, all Fs < .16, all ps > .34.

Super-Learning. As shown in Figure 17, participants showed enhanced, super-learning to Y and W that was facilitated by their training in compound with a previously predictive inhibitor and excitor respectively. As shown in the upper panel of Figure 17, participants rated Y ($M = 44.18$, $SD = 23.77$) as significantly more excitatory than S ($M = 26.01$, $SD = 28.82$), $F(1,32) = 5.11$, $p = .031$, $\eta_p^2 = .14$, 95% CI (.00, .35). No interactions were significant, all Fs < .88, all ps > .35. As shown in the lower panel of Figure 17, participants rated W ($M = -51.38$, $SD = 27.60$) as significantly more inhibitory than Z ($M = -28.34$, $SD = 29.98$), $F(1,32) = 12.36$, $p = .001$, $\eta_p^2 = .28$, 95% CI (.05, .48). No interactions were significant, all Fs < 2.33, all ps > .89.

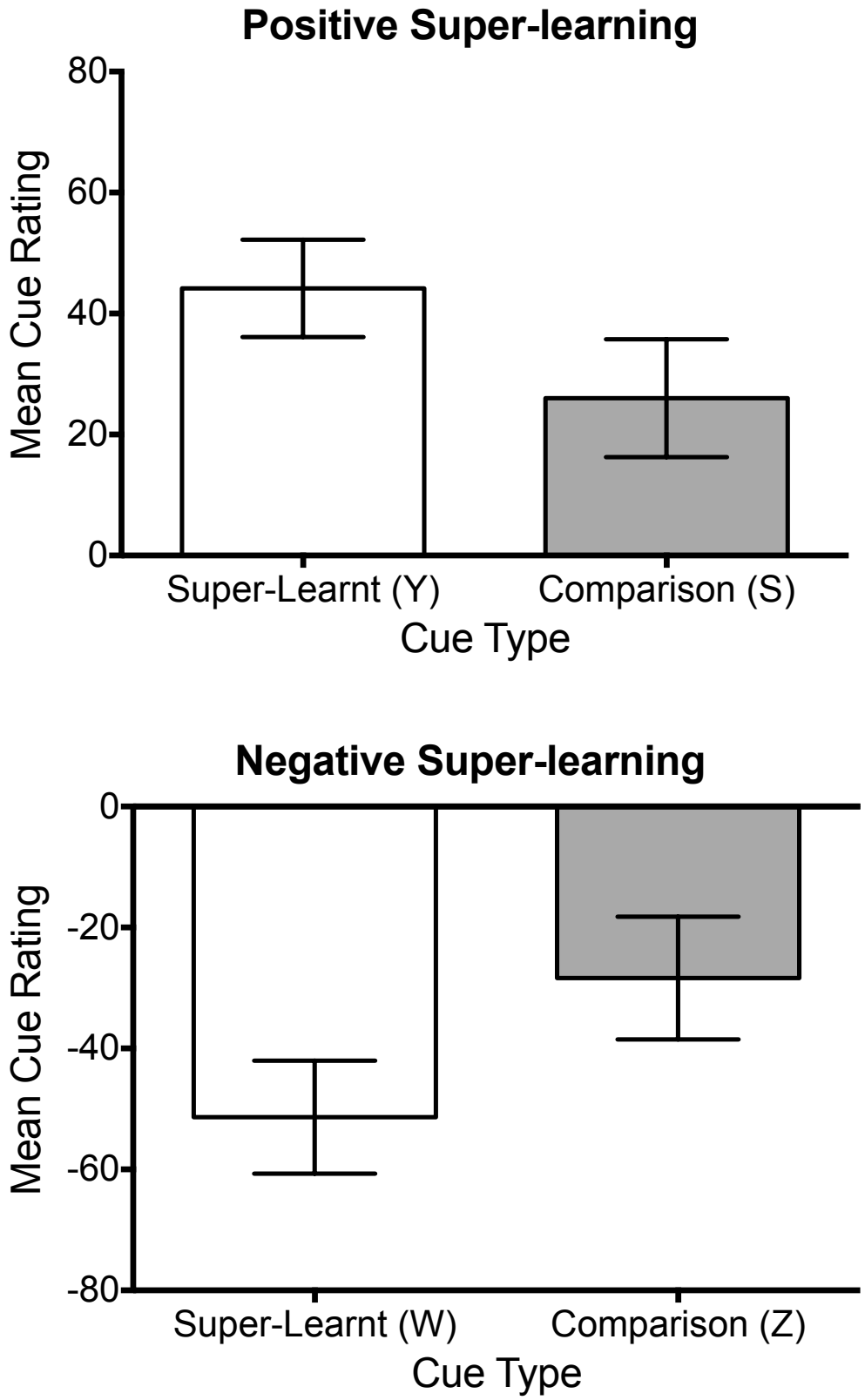


Figure 17. Participants showed super-learning to cues Y and W as compared to cues S and Z respectively. Error bars represent 95% confidence intervals.

3.6.3.2. Relationship Between Psychopathy and Dependent Variables

The zero-order correlations between overall psychopathy score, the factors of SCI, FD and C, and the dependent variables are shown in Table 15. FD was significantly related to shift ratings to the inhibitor (A; $r = .40$, $p = .016$), and to positive super-learning ($r = -.60$, $p < .001$). All other dependent variables were not significantly related to psychopathy.

Table 15

Experiment 5 Zero-Order Correlations

	Total PPI-R	**SCI**	**FD**	**C**
Total PPI-R	-			
SCI	**.541****	-		
FD	**.732*****	-.034	-	
C	.135	-.277	.021	-
A Shift Errors	-.061	-.305	.160	-.058
B Shift Errors	.100	-.049	.165	.051
A Shift Ratings	**.457****	.141	**.398***	.205
B Shift Ratings	-.276	-.110	-.184	-.259
C Shift Ratings	.128	.060	.099	-.067
D Shift Ratings	-.002	-.155	.177	-.037
Positive SL	**-.419***	.037	**-.596*****	.072
Negative SL	-.099	-.041	-.053	.131

Note: Total PPI-R, SCI, FD and C represent the *t*-scores for each variable. Positive SL = Positive Super-Learning Ratings; Negative SL = Negative Super-Learning Ratings.

*** $p < .001$; ** $p < .01$; * $p < .05$.

As in Experiments 3 and 4, where significant correlations were found between psychopathic traits and inhibitory, but not excitatory, learning, the differences in these correlations was tested using Fisher's *r*-to-*z* transformation and a *z*-test (I. A. Lee & Preacher, 2013). This showed that the difference in the correlation between overall psychopathy traits and excitatory and inhibitory learning was not significant for the shift learning, but was significant for the second-order 'super-learning' effects (z = -2.24, p = .025). Likewise, the difference in correlation between the factor of FD and excitatory and inhibitory learning was not significant for the shift learning, but was significant for the second-order 'super-learning' effects (z = -3.03, p = .002).

As shown in Figure 18, participants with higher levels of FD showed greater shift learning to the inhibitory cue [A; β = .83, 95% CI (.13, 1.54), $t(35)$ = 2.41, p = .022, η_p^2 = .16, 95% CI (.01, .35)], and weaker super-learning [β = -1.78, 95% CI (-2.74, -.79), $t(35)$ = -3.71, p = .001, η_p^2 = .29, 95% CI (.06, .49)]. Only this direct effect of Fearless Dominance on super-learning was significant: the indirect mediated effects were not significant, all βs < .06, all 95% CIs (< -.13, > .01). FD traits were not directly or indirectly predictive of negative super-learning, all βs < .28, all 95% CIs (< -.02, > .26).

SCI traits were not directly or indirectly predictive of either positive super-learning [all βs < .33, all 95% CIs (< -.04, > .04)], or negative super-learning [all βs < .12, all 95% CIs (< -.30, > .02)]. Being male did weakly predict reduced negative super-learning, [β = 31.31, 95% CI (.98, 61.65), $t(35)$ = 2.12, p = .044, η_p^2 = .12, 95% CI (.00, .32)], however the 95% confidence intervals about the effect size crossed zero, suggesting that sex may not predict differences in negative super-learning.

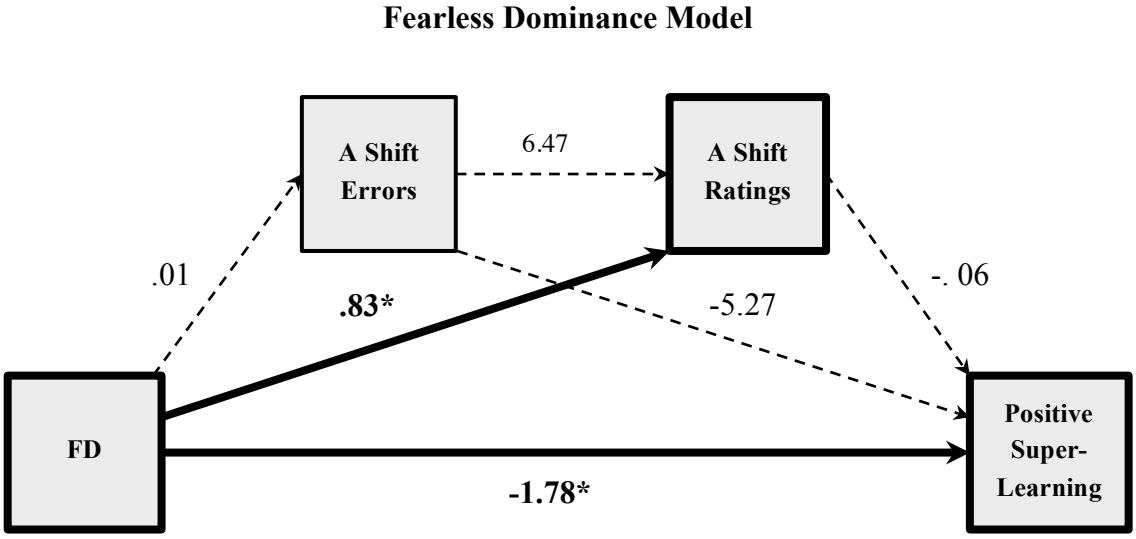


Figure 18. The numbers shown represent β values, signalling the strength of a relationship between two variables. FD directly predicted shift ratings of the inhibitory cue (A) and positive super-learning. The indirect, mediated relationships between FD and positive super-learning were not significant.

* $p < .05$

The simultaneous multiple linear regression analysis showed that psychopathy factors were not significantly predictive of a shift in test cue C, $R^2_{adjusted} = .053$, $F(4,35) = .43$, $p = .784$, $\eta_p^2 = .09$, 95% CI (.00, .13) all βs < .22, all ps > .27, or test cue D, $R^2_{adjusted} = .070$, $F(4,35) = .58$, $p = .678$, $\eta_p^2 = .01$, 95% CI (.00, .01) all βs < .19, all ps > .30.

3.6.3.3. *Post Hoc* Analysis of the Inhibitory Cue

Given the significant relationship between shift learning to the inhibitor (A) and FD traits, *post hoc* comparison of participant ratings of A and V in the acquisition phase was carried out as in Experiments 3 and 4. A 2 (cue: A vs. V) x 2 (flavour counterbalancing) x 2 (sex) mixed model ANOVA confirmed that ratings of A (M = -38.21, SD = 25.74) were significantly lower than the ratings of V (M = -2.68, SD =

25.07), $F(1, 32) = 8.96$, $p = .005$, $\eta_p^2 = .22$, 95% CI (.02, .43). A simultaneous multiple linear regression analysis including FD, SCI, sex and flavour counterbalancing was not significant, all βs < .14, all ps > .49.

3.6.4. Discussion

This study showed that psychopathic traits are not related to a general inability to distinguish predictive and non-predictive stimuli, or to generally rapid development of excitatory associations. In addition, the results of Experiments 3 and 4 were broadly replicated, showing that psychopathic traits predict a difference in updating learning about inhibitory cues, but not excitatory cues, something that does not appear to be due to poor excitatory association formation.

These findings are important, for not only do they broadly replicate the results of the previous studies, they also help to establish the specificity of the difference seen in learning in relation to psychopathic traits. On the basis of Experiments 2-4, it was not possible to rule out the suggestion that psychopathic traits were simply associated with more rapid acquisition of excitatory associations, which would lead to greater shift to the inhibitory cue. If this were true, psychopathic traits predicted increased learning about D, the neutral cue that was re-trained as an excitatory cue in the shift phase. However, in the present sample psychopathic traits were not predictive of differences in relation to learning about either of the previously non-predictive cues. This indicates that the learning differences seen in relation to psychopathic traits are specific to inhibitors, and the vulnerability surrounds these associations.

In addition, these results suggest that psychopathic traits are not associated with difficulties in forming excitatory associations, which was a possible explanation for the observed inhibitory learning differences given the dependence of inhibitory association

development on excitatory cues (Rescorla, 1969). Psychopathic traits did not predict differences in the ability to distinguish a neutral and an excitatory cue in the acquisition phase, and nor did they predict differences in shifting to acquire an excitatory association to a previously neutral cue. As such, it appears that inhibitory learning differences were not driven by weaker excitatory learning. However, the excitatory cue that was tested was trained in compound (BV+), and so it is for future studies to clarify this point by assessing simple excitatory learning to an element (e.g., B+).

3.7. Summary of Chapter 3 Results

Shift Errors. Overall, Experiments 2-5 showed that in community samples, psychopathic traits are not related to difficulties in learning when to respond and when to withhold responding. In Experiments 4 and 5 there were no differences in errors to the test cues across the phases. In Experiments 2 and 3 there was an increased number of response errors to the excitatory test cue in the acquisition phase. Whilst in Experiment 2 this was unrelated to psychopathic traits, in Experiment 3 this difference was positively predicted by FD, suggesting that high FD trait individuals found it particularly difficult to learn to respond to the excitatory cue in the acquisition phase, and easier to learn to withhold responding to the cue in the shift phase. Psychopathic traits were also not related to SSRT, which was measured in Experiment 2.

Shift Ratings. Experiments 3-5 showed that psychopathic traits (FD and SCI) are predictive of enhanced inhibitory shift learning but comparable excitatory shift learning. Analysis of participants' shift learning to neutral cues retrained as either excitatory or inhibitory showed that this was also not predicted by psychopathic traits. Therefore, the observed differences do not appear to reflect the more rapid, or overdeveloped acquisition of excitatory associations, and instead suggest a problem unique to the development and shift of inhibitory associations.

Another interesting result that emerged in Experiments 3 and 4 was the relationship between sex and the flexible updating of the previously trained excitor when the contingency was changed. It was found that female participants showed a greater ability than their male counterparts to learn that a cue that had previously signalled reward was now less likely to do so. This difference was not seen in Experiment 2 where the magnitude of change was larger, indicating that the sex-specific difference appears to concern sensitivity to subtle reductions in a stimulus' prediction of reward. This

difference was also not seen in Experiment 5; however, this may have been due to the complexity of the design.

Super-Learning. Participants appear to have shown super-learning, where a previously trained cue facilitates a novel cue's acquisition of an opposing association. This was shown in the acquisition of both enhanced excitatory and enhanced inhibitory associations. Whilst the facilitation of enhanced excitatory associations is well documented in animals and humans (Rescorla, 1990; B. A. Williams & McDevitt, 2002), the facilitation of enhanced inhibitory associations is, to the best of my knowledge, a relatively new phenomenon. It serves to add to the literature suggesting that inhibitory and excitatory processes are symmetrical opposites (Rescorla, 1969).

The extent of the positive super-learning was predicted by psychopathic traits in Experiments 3 and 5, but not in Experiments 2 and 4. I suggested that the comparison cue design might be vulnerable to overshadowing or blocking (Kamin, 1969) in Experiment 3 and so the designs of Experiments 4 and 5 were altered to reduce the associability of the cues trained in compound with the super-learnt comparison cues (i.e., C and D), whilst retaining the matched design in the shift phase. In Experiment 4, cues C and D were trained as non-predictive elements and in compound with a predictive stimulus (i.e., C+/-, CV-; D+/-; DX+), and so it was predicted that learning about C and D would be blocked by the more predictive cues with which they were trained (i.e., V and X). In contrast, in Experiment 5 cues C and D were only trained as non-predictive (i.e., C+/-, CT+/-; D+/-, DU+/-). Therefore, it is possible that participants were better able to acquire associative strength in the shift phase in Experiment 5, where the cues were only pre-exposed as non-predictive, than they were in Experiment 4, where the cues were pre-exposed as non-predictive in the presence of a predictive cue, the latter possibly leading to greater reduced associability. This would mean that the comparison cues were

more likely to be blocked or overshadowed in Experiment 5 than they were in Experiment 4 as the cues with which they were trained were more easily learnt about. This pattern of results is broadly consistent with the suggestion that because of a narrow processing bottleneck (Hamilton et al., 2015), high psychopathy trait individuals are more robust to overshadowing or blocking.

As discussed in relation to Experiments 3 and 4, the observed super-learning results may have been driven by learning in the shift phase, resulting in reduced learning to the comparison cues, as opposed to learning in the acquisition phase subsequently enhancing learning to the super-learnt cues. However, were the apparent super-learning effects due to overshadowing or blocking, the prediction would follow that this super-learning effect would be reduced in Experiment 5 as compared to Experiment 3. This is because in Experiment 5 the cues with which the comparison cues were trained were pre-exposed as non-predictive (CT+/-, DU+/-), meaning that their alpha values will decrease and so subsequent learning about these cues will be impaired (Mackintosh, 1975). This means that these cues should be less able to overshadow or block the super-learning comparison cues (Z and S) in the shift phase when they are presented together with the pre-exposed cues in compound (CZ-, DS+). As such, the super-learning comparison cues will be able to acquire greater associative strength, and so the super-learning effect will be reduced compared to Experiment 3 in which this pre-exposure of C and D did not occur.

However, when examining the difference in ratings between the super-learning and comparison cues, this prediction was not supported. For positive super-learning: the difference in Experiment 3 (M = 10.95) is in fact smaller than that of Experiment 5 (M = 18.17). Likewise, for negative super-learning: the difference in Experiment 3 (M = -9.56) is in fact smaller than that of Experiment 5 (M = -23.04). These comparisons have been

made *post hoc* and statistical analysis has not been run. However, there is no evidence that the super-learning is weaker in Experiment 5 than in Experiment 3, an observation inconsistent with the suggestion that the reported super-learning effects were driven solely by the restricted learning about the comparison cues. The comparison cue design in Experiments 3-5 is, nonetheless, not ideal for enabling a firm conclusion to be made regarding super-learning. Whilst the between-experiment comparisons made here may go some way to clarify the nature of these results, it remains for future research to determine whether the super-learning effect can be observed using stricter comparison cues. As such, all discussion of 'super-learning' should be interpreted with the appropriate caveat that this effect requires further clarification in order to definitively establish whether super-learning is occurring.

3.8. General Discussion

Psychopathy only predicted response errors in one of the four experiments presented in this chapter, and in fact the results suggest that high psychopathy individuals found it easier to withhold their responding to a cue previously signalling the need to make a response. This poses a challenge to theories designed to account for psychopathic impulsivity. Both attentional and outcome processing accounts are based on evidence of altered responding in psychopathic individuals (e.g., Blair et al., 2004; Budhani et al., 2006; Newman & Schmitt, 1998; Patterson & Newman, 1993). However, as discussed in Chapter 1, studies supporting these accounts did not separately assess behavioural and cognitive contributions to impulsivity. Therefore, it appears that when these two processes are separately tested it is the explicit learning component that is impaired, and not the simple motoric aspect of the task. This assertion is supported by psychopathic traits also not predicting performance on the stop-signal task, an independent measure of

participants' behavioural impulsivity. As such, it is argued that it is the complex learning component that has driven response differences in previous research, something that was prevented here by the use of the novel go/nogo task, and something that is not tested in a stop-signal task where a rule is given and so learning is simpler and more response orientated.

The novel task used in these experiments contained a balanced number of trials requiring participants to make or withhold a response. As was discussed in Chapter 1, the failure to use such a balance can bias participants' responding (Kiehl et al., 2000). This is of particular importance to attentional theories as an imbalance of trials may induce additional attentional demands required for withholding responding to an inhibitory cue. Therefore, the present studies are consistent with other research that has used a balanced number of go and nogo trials and found that psychopathic traits are not related to behavioural impulsivity problems (e.g., Kiehl et al., 2000; Sadeh & Verona, 2008).

Psychopathic traits were predictive of inhibitory learning, specifically weak inhibitory learning. This difference was apparent in direct tests (shift) and indirect tests (e.g., super-learning) of the inhibitor. Given the specificity of this difference in that it only affected inhibitors and not excitors, it does not appear that psychopathic traits are associated with general learning or contingency detection problems. In each experiment participants demonstrated that they were capable of learning about cues, a process dependent on learning the contingencies between stimuli and outcomes. If, as outcome processing accounts suggest (e.g., Blair, 2005), there was a general problem in psychopathic individuals' ability to process outcomes then more general learning impairments would have been seen. Furthermore, it would not have been possible for participants to update their learning about the excitatory and inhibitory cues in the shift phase. As these learning and shift impairments were not seen, it suggests that participants

did not struggle to process the outcome. The differences in relation to psychopathic traits also do not appear to result from overdeveloped excitatory associations, which might have facilitated rapid shift. Were this to have been the case, psychopathic traits would have been expected to predict advanced excitatory acquisition to the previously non-predictive cue in Experiment 5.

Instead, these results can be accounted for by the difficulty in determining a cue's relative validity, another process dependent on the integration of information (Mackintosh, 1975). Mackintosh suggested that determining a cue's relative validity is integral to its associability (α), and so how much attention is directed to it and how easily it is learnt about. If psychopathic traits predict a difficulty in integrating information (Hamilton et al., 2015; Hoppenbrouwers et al., 2015), then α may not adjust to reflect the cue's relative validity with the same degree of sensitivity that is seen in non-psychopathic individuals. Therefore, in the shift phase low psychopathic trait individuals' α may be reduced as the relative validity of test cue A is also reduced, meaning that learning about test cue A will be slowed. In contrast, high psychopathic trait individuals' α may remain unchanged, meaning that learning about test cue A will be facilitated throughout the shift phase. By interpreting psychopathic learning differences through Mackintosh's attention model it is possible that a mechanistic account of psychopathic learning and behaviour can be offered.

There is some evidence that this inability to integrate information affects the initial development of inhibitory associations. Individuals with low levels of psychopathic traits are able to integrate the information from the different stimulus-outcome pairings, and so acquire stronger inhibitory associations which are then less easily changed (Mackintosh, 1975; Rescorla & Wagner, 1972; Rescorla, 1969). In contrast, high psychopathy trait individuals appear to form an 'inhibitory association' that is comparable to that formed

when a cue is simply paired with the absence of the outcome. However, this effect was only seen in Experiment 3, and not in Experiments 4 and 5. This suggests that the primary difference responsible for the altered learning in psychopathic individuals relates to their ability to adjust a cue's associability, and that the difficulty integrating information which causes this effect may in some instances affect the initial acquisition of inhibitory association. As the conditions under which such initial learning differences occur were not tested, it is not possible to make further conclusions regarding this initial acquisition of inhibition.

With respect to the apparent super-learning observed here, it has been discussed that there are limitations in the designs used to train the comparison cues in Experiments 3-5. As previously explained, this occurred in part as a consequence of matching the shift phase training for the super-learnt and comparison cues. Experiments 4 and 5 used alternative designs that retained the same shift phase training, but pre-exposed certain cues in the acquisition phase to try and make the comparison more stringent. The details of these design changes and their effects are outlined in the results summary (3.7.). However, whether or not this is considered as an example of 'super-learning', these experiments do show that people learn differently about novel cues depending on whether they are trained in compound with a stimulus predictive of the same or an opposing outcome. Therefore, this still acts as an indirect, second-order test of learning, and the differences observed in relation to psychopathic traits are still suggestive of an altered ability to combine information across stimuli.

3.9. Conclusions

Using a novel go/nogo paradigm this series of experiments has shown that psychopathic traits are predictive of altered inhibitory learning, demonstrated through

enhanced inhibitory shift learning. Participants' excitatory learning and their ability to respond and withhold responses to cues were not (generally) predicted by psychopathic traits. I have argued that an outcome processing account cannot explain the specificity of these findings, but that an altered associability account can. Specifically, these results suggest that psychopathic traits are associated with a difficulty in integrating information, a process necessary in the development of inhibitory associations and in the adjustment of a cue's associability. As such, these findings support the interpretation of psychopathy through an attentional-associative account.

CHAPTER 4

The Effect of Cognitive Load on Shift Learning

(Experiment 6)

4.1. Introduction

On the basis of the conclusions drawn in the previous chapter, predictions can be made about the circumstances under which inhibitory learning differences in psychopathy may or may not occur. For instance, it has been suggested here and elsewhere that there is an attentional component to psychopathy (Gorenstein & Newman, 1980; Newman & Baskin-Sommers, 2011), and that this restricts the processing capacity of high psychopathy trait individuals (Baskin-Sommers et al., 2013; Hamilton et al., 2015). This reduced processing capacity, or 'processing bottleneck', is a central feature of the Impaired Integration Theory, and led Hamilton and colleagues to argue that this causes processing in psychopathic individuals to be akin to that seen in low psychopathy trait individuals under high cognitive load conditions. Though the causes are different, one due to a narrow bottleneck, the other due to increased load, both result in the same reduced processing of the available information, and altered input-to-output ratio, as compared to processing in low psychopathy individuals under normal cognitive load conditions. A representation of these differences can be seen in Figure 19.

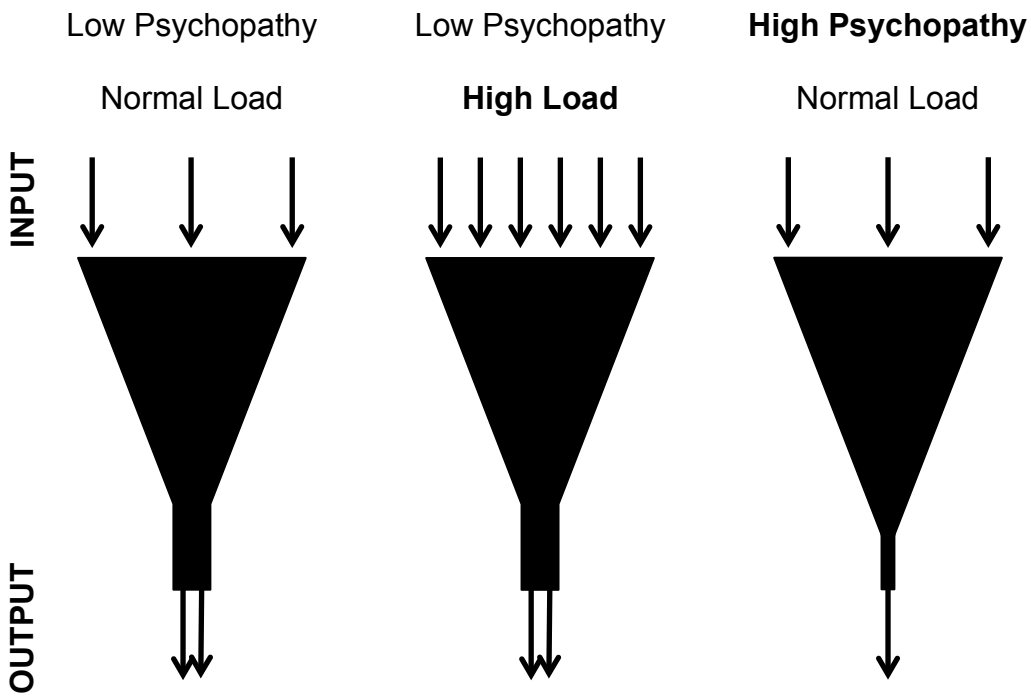


Figure 19. Load = cognitive load. Processing bottlenecks under three conditions as suggested by the Impaired Integration Theory (Hamilton et al., 2015): the left bottleneck shows processing in low psychopathy individuals under normal cognitive load conditions; the middle bottleneck shows processing in low psychopathy individuals under high cognitive load conditions; and the right bottleneck shows processing in high psychopathy individuals under normal cognitive load conditions. The ratio of input-to-output in the high cognitive load and the high psychopathy conditions is worse (i.e., less of the available information is processed) than it is in the normal cognitive load condition.

Were participants given an additional load task to perform alongside the go/nogo task, it was predicted that individuals with low psychopathic traits would be induced to shift their learning like those with high psychopathic have consistently been shown to do. I argue that this would occur because the additional cognitive load would limit

individuals' processing capacity, and so their ability to integrate information from different stimulus-outcome pairings, a process necessary when determining the relative validity of a stimulus and so how easily it is learnt about (Mackintosh, 1975).

In order to test this prediction, Experiment 6 required participants to perform an additional cognitive load task concurrently with the go/nogo task. Baddeley & Hitch's (1974) Theory of Working Memory suggested that there are three components of working memory: the phonological loop; the visuo-spatial sketchpad; and, the central executive. However, more recent evidence suggests that these are less distinct than initially thought (Baddeley, Chincotta, & Adlam, 2001). The central executive is believed to control overall attention, whilst the phonological loop was thought to specifically relate to verbal short-term memory, and the visuo-spatial sketchpad specifically to visual short-term memory.

Baddeley & Hitch's (1974) theory suggested that if cognitive load were high, then some of the working memory capacity would be occupied, and this would influence task performance. For example, previous research has shown that continuously repeating a word increases cognitive load and prevents rehearsal using the phonological loop, which can impact on go/nogo task performance (Baddeley et al., 2001; Baddeley, Lewis, & Vallar, 1984; Dunbarb & Sussman, 1995; Tullett & Inzlicht, 2010). Therefore, in this experiment participants were required to repeat the word 'the' during the go/nogo task, a word that has previously been used in such an articulatory suppression task (Baddeley, Chincotta, & Adlam, 2001, Study 6).

4.2. Method

4.2.1. Participants and Design

Forty-seven individuals (21 male, 26 female; age $M = 24.15$ years, $SD = 2.92$)

participated, using the same design as in Experiment 3. The same number of participants was tested as in Experiment 3 in order to allow a between-experiment comparison. Participant recruitment and eligibility criteria and ethical approval were the same as in previous experiments (3.2.2.).

4.2.2. Materials

Psychopathic Personality Inventory (Revised; PPI-R; Lilienfeld & Widows, 2005). The internal consistency of the PPI-R was good (α = .89). The means and standard deviations for the participants are shown in Table 16.

Table 16

Experiment 6 Participant Psychopathy *T*-Scores

	M	*SD*
Total PPI-R	49.55	10.42
Self-Centred Impulsivity	53.64	10.04
Fearless Dominance	45.72	10.48
Coldheartedness	47.34	10.04

Note: N = 47. The mean *t*-score is 50.

Shift Go/NoGo Task. This was the same as in Experiment 3, with the addition that participants were required to continually repeat the word 'the' out loud throughout the go/nogo task. As explained in the introduction, this additional task has previously been shown effective at increasing cognitive load when used in conjunction with a go/nogo task (e.g., Tullett & Inzlicht, 2010).

4.2.3. Statistical Analysis

Statistical analysis was the same as in Experiment 3, in which the dependent

measures were: shift go/nogo errors (A and B); shift ratings (A and B); and super-learning ratings (Y and W). These were analysed for either a main effect of phase or cue type as in the previous experiments, using mixed model analyses of variance, which included counterbalancing condition and sex as between-subject variables. In addition, zero-order correlations between the overall psychopathy score, the three-psychopathy factors (Self-Centred Impulsivity, SCI; Fearless Dominance, FD; and, Coldheartedness, C) and the dependent variables were calculated.

Multiple mediation analyses with mediators operating in serial were conducted to test whether either FD or SCI predicted super-learning, and whether these relationships were mediated by shift errors or shift ratings. As before: counterbalancing condition and sex were also included in this analysis; an alpha of .05 was used; *t*-tests were only conducted following significant ANOVA results and were corrected using Bonferroni; and, 95% confidence intervals were calculated for any ANOVA, *t*-test or mediation analysis where $p < .05$. The confidence intervals were reported to 2 decimal places, and any confidence interval that included or crossed zero is explicitly flagged, with the implications of this discussed.

4.3. Results

Overall, psychopathic traits were not related to response errors, a finding consistent with the previous experiments presented here. As shown in Figure 21, psychopathic traits were also not significantly related to the extent of shift learning to either cue A, the strong inhibitor that was then trained as a weak inhibitor, or cue B, the strong excitor that was then trained as a weak excitor. Consistent with the extensive shift to test cues A and B, participants did not show either positive or negative super-learning, the second-order effects that rely on these cues.

4.3.1. Overall Learning Effects

Initial Cue Learning (Acquisition Phase). As shown in Figure 20, participants rated the inhibitory cue (A; *M* = -31.24, *SD* = 25.58) as significantly more negative than the excitatory cue (B; *M* = 39.30, *SD* = 21.98), $F(1, 43) = 169.99$, $p < .001$, $\eta_p^2 = .80$, 95% CI (.67, .86). No interactions were significant, all *F*s < 1.98, all *p*s > .16.



Figure 20. Participants rated the inhibitory cue (A) as significantly more negative than the excitatory cue (B) in the acquisition phase. Error bars represent 95% confidence intervals.

Shift Errors. There was no overall difference in errors to test cue A in the acquisition (*M* = .49, *SD* = .83) and shift (*M* = .79, *SD* = 1.35) phases, all *F*s < 1.99, all *p*s > .17. However, for test cue B participants made significantly more errors in the acquisition phase (*M* = 1.77, *SD* = 1.97) than they did in the shift phase (*M* = .26, *SD* =

.71), $F(1,43) = 32.21$, $p < .001$, $\eta_p^2 = .43$, 95% CI (.20, .58). No interactions were significant, all Fs < .37, all ps > .59.

Shift Ratings. As shown in Figure 21, overall participants shifted their learning to test cue A (acquisition: $M = -31.33$ $SD = 23.46$; shift: $M = -15.41$, $SD = 25.37$) the previously predictive inhibitor, $F(1,43) = 12.24$, $p < .001$, $\eta_p^2 = .22$, 95% CI (.04, .41). No interactions were significant, all Fs < .34, all ps > .76. Likewise, participants shifted their learning to test cue B (acquisition: $M = 39.30$, $SD = 21.98$; shift: $M = 26.79$, $SD = 23.89$), the previously predictive excitor, $F(1,43) = 10.13$, $p = .003$, $\eta_p^2 = .19$, 95% CI (.03, .38). No interactions were significant, all Fs < 1.49, all ps > .22.

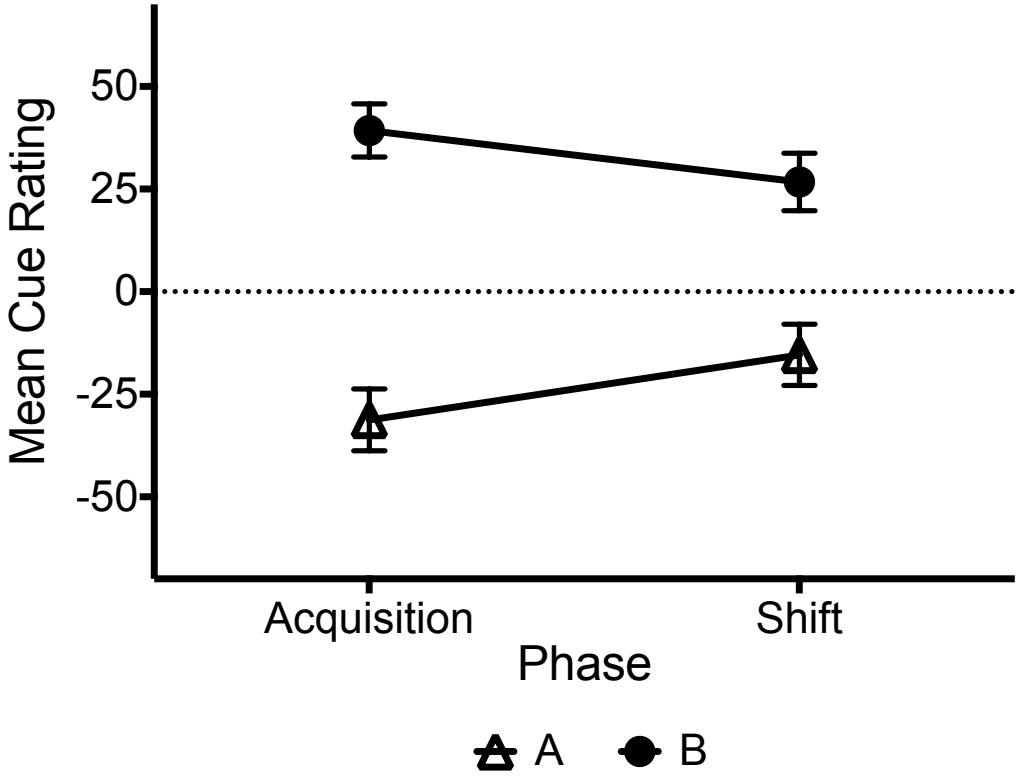


Figure 21. Participants showed significant shift to the inhibitory (A) and excitatory (B) test cues. Error bars represent 95% confidence intervals.

Super-Learning. When under additional cognitive load, participants did not show enhanced learning to Y (M = 40.78, SD = 23.65) as compared to S (M = 42.72, SD = 22.73), all Fs < 1.62, all ps > .20. Nor did they show enhanced learning to W (M = -43.05, SD = 18.70) as compared to Z (M = -44.30, SD = 24.04), all Fs < 1.12, all ps > .29.

4.3.2. Relationship Between Psychopathy and Dependent Variables

The zero-order correlations between overall psychopathy score, the factors of SCI, FD and C, and the dependent variables are shown in Table 17. Once again, each of the three PPI-R factors was strongly related to the overall PPI-R score. Negative super-learning was significantly related to Coldheartedness (r = -.31, p = .033). All other dependent variables were not significantly related to psychopathy.

There was a small direct effect of SCI on A's shift ratings [β = .90, 95% CI (.02, 1.78, t(46) = 2.05, p = .046, η_p^2 = .08, 95% CI (.00, .25)]. However, as the 95% confidence interval about the effect size crossed through zero, it suggested that there may be no effect of SCI on shift learning to the inhibitor. No other direct or indirect effects between SCI and positive super-learning were significant [all βs < .40, all 95% CIs (< -.03, > .01)], nor were the direct or indirect effects between SCI and negative super-learning significant [all βs < .21, all 95% CIs (< -.07, > .02)]. None of the direct or indirect effects between FD and either positive super-learning [all βs < .55, all 95% CIs (< -.02, > .04)] or negative super-learning [all βs < .36, all 95% CIs (< -.01, > .09)] were significant.

Table 17

Experiment 6 Zero-Order Correlations

	Total PPI-R	SCI	FD	C
Total PPI-R	-			
SCI	**.787*** **	-		
FD	**.721*** **	.177	-	
C	**.393** **	.091	.275	-
A Shift Errors	-.010	-.079	.042	.126
B Shift Errors	-.014	.058	-.089	.037
A Shift Ratings	.156	.286	-.024	-.126
B Shift Ratings	.120	.000	.255	-.123
Positive SL	.008	-.071	.037	-.127
Negative SL	-.099	-.094	-.041	**-.312***

Note: Total PPI-R, SCI, FD and C represent the *t*-scores for each variable. Positive SL = Positive Super-Learning Ratings; Negative SL = Negative Super-Learning Ratings.

*** $p < .001$; ** $p < .01$; * $p < .05$.

4.3.3. *Post-Hoc* Comparison of Experiments 3 and 6

Given that the design in Experiments 3 and 6 were identical, save for the inclusion of an additional cognitive load task in Experiment 6, a between-experiments analysis was done in order to test the effect of cognitive load. This between-experiment analysis was done exclusively for the shift in inhibitory learning (about cue A) as this is where the principal psychopathy differences were identified in Experiment 3 and was the variable hypothesised to be most affected by cognitive load. Moderation analysis was carried out

to test whether cognitive load moderated the effect of FD traits on inhibitory shift learning. These relationships were tested using parametric bootstrapping procedure (PROCESS; Preacher & Hayes, 2008) which estimates the effect of the moderator variable, and as with the mediation analysis, was based on 1000 bootstrap resamples with 95% bias corrected confidence intervals. Sex and flavour counterbalancing condition were included as covariates in the analysis. Though the overall model only approached significance, $F(5,88) = 2.06$, $p = .078$, cognitive load had a significant effect on the relationship between FD traits and inhibitory shift [$t(93) = 1.91$, $p = .053$, $\eta_p^2 = .04$, 95% CI (.00, .14)], with the conditional effect of FD on inhibitory shift significant in the low load condition [$t(46) = 2.08$, $p = .041$, $\eta_p^2 = .05$, 95% CI (.00, .16)], but not the high load condition.

4.4. Discussion

The results of Experiment 6 show that psychopathic traits do not predict inhibitory shift learning under high cognitive load conditions. The relationship between psychopathic traits and enhanced shift learning about an inhibitor, which is shown in Experiments 3-5, is not seen here. This appears to suggest that when artificially placed under high cognitive load, individuals with low psychopathic traits perform more similarly to those with high psychopathic traits. As such, these results offer preliminary support for the account generated in Chapter 3, according to which psychopathic individuals have reduced perceptual processing, which can be likened to the reduced perceptual processing seen in high cognitive load conditions in low psychopathy trait individuals.

The absence of second-order effects of the excitatory and inhibitory shift cues provides further evidence suggestive of the low psychopathy trait individuals acting more

like the high psychopathy trait individuals. Experiments 1-5 have shown that, in general, participants distinguish between novel cues trained in compound with a cue predictive of the same or of an opposing outcome. This has been suggested to indicate significant main effects of 'super-learning', though given design limitations the definitiveness of these results has been questioned, and it has been acknowledged that further research is required to clarify the nature of these findings. Whether-or-not these results can be interpreted as showing super-learning, they do indicate that participants differentiate between cues paired with the outcome which are trained in the presence of a cue predictive of the same or of the opposing outcome.

Psychopathic traits have predicted differences in these second-order effects (Experiments 3 and 5) in a manner suggestive of increased resistance to interference by other cues, resulting in weaker distinction between training conditions, and so reduced super-learning. Therefore, the absence of a main effect of super-learning in the current study suggests that the low psychopathy trait individuals have performed more similarly to the high psychopathy trait individuals, and not the other way around. If the high psychopathy trait individuals had been induced to act more like their low psychopathy trait counterparts, then second-order learning effects, such as super-learning or blocking, would have been expected.

However, cognitive load was not an experimental manipulation, and the comparisons made here are done *post hoc* across Experiments 3 and 6. As such, whilst these results may suggest that cognitive load influences contingency monitoring in low psychopathy trait individuals, they cannot be regarded as definitive evidence for this assertion. In order to make such a conclusion, a further study, which manipulated cognitive load as a between-subjects variable, would need to be conducted. Nonetheless, as the present experiment was identical in design to Experiment 3, the disparity in results

suggests that cognitive load does affect shift learning in relation to psychopathic traits, and that this is most likely because under high load conditions low psychopathy trait individuals act more like high psychopathy trait individuals.

4.5. Conclusions

This experiment showed that under conditions of high cognitive load psychopathic traits do not predict shift learning. These results are consistent with the Impaired Integration Theory (Hamilton et al., 2015), which suggests psychopaths operate under conditions akin to high load conditions in non-psychopathic individuals. This is also consistent with the associability account, which suggests that a high load state impairs psychopathic individuals' ability to integrate information in order to determine a cue's relative validity, and so its associability.

CHAPTER 5

Contextual Processing and Perceptions of Control

(Experiment 7)

5.1. Introduction

In Chapters 2-4 I have explored a novel means of assessing learning and contingency monitoring in relation to psychopathic traits. The findings of these studies suggest that psychopathic traits predict increased shift learning about inhibitory cues but not excitatory cues, and differences in the way inhibitory cues influence learning about other cues (i.e., second-order effects). Using an associative-attentional account based on Mackintosh's (1975) theory, these differences have been interpreted as a failure to integrate information in order to establish a cue's relatively validity, and so its associability. The Impaired Integration Theory (Hamilton et al., 2015) suggests that this altered associability results from shallow information processing, exacerbated by the resulting failure to process peripheral stimuli (Baskin-Sommers et al., 2013).

If, as has been suggested, psychopathic traits are associated with overly focused attention at the expense of processing peripheral information (e.g., Hamilton et al., 2015; Wolf et al., 2012), then it follows that the context will only be superficially learnt about, if indeed it is learnt about at all (Hoppenbrouwers et al., 2015). The shift go/nogo task has provided a suitable means to assess contingency monitoring of stimulus-response-outcome associations, and the indirect affect of integrating contextual information. However, it does not provide a direct assessment of participants' processing and learning

about the context in which learning takes place, or how this contextual information is integrated to inform learning about other, more central stimuli. Therefore, whilst in the previous experiments I sought to change S-R-O contingencies between experimental phases, here I examined the pure effect of contingency learning within a session using a variable or 'shifting' contingency (zero contingency) and a predictive 'excitatory' contingency (positive contingency). Using a probabilistic causal learning task (Dickinson, Shanks, & Evenden, 1984; Msetfi, Murphy, Simpson, & Kornbrot, 2005; Shanks, 2007), I tested how perceptions and ratings of control differed in these controllable and uncontrollable conditions, and whether this was related to psychopathic traits.

In probabilistic causal learning tasks participants learn about the relationship between an action (the self), the environment (the context) and an outcome in order to test their perceptions of causal control, and ability to process and form associations with the context (Msetfi, Murphy, & Simpson, 2007; Msetfi, Wade, & Murphy, 2013). Much like go/nogo tasks, participants perform an action or not and the outcome occurs probabilistically after both events. However, unlike go/nogo tasks, participants make ratings about the strength of the perceived association between their actions and the outcome, and importantly, the perceived association between the context and the outcome. The assessment of the context is important as personal control is thought to be evaluated in contrast to the context-outcome relationship that might compete for causal control (Byrom et al., 2015). This is particularly relevant because of evidence for impaired contextual processing in psychopathy (Brazil et al., 2011; Hoppenbrouwers et al., 2015; Wolf et al., 2012).

Learning that one's responding is more effective than the constellation of other possible causes, nominally referred to as the context, has been characterised as involving

a process of overshadowing, whereby one possible cause is learnt less well in the presence of another cue (Darredeau, Baetu, Baker, & Murphy, 2009; see also Rescorla, 1971a). Similarly learning that the context has more effective causal control might be expected to involve the overshadowing of the action by the context. The characteristic self-focused approach in psychopathy may disrupt this process of overshadowing: if psychopathic individuals' attentional allocation to stimuli is imbalanced towards central stimuli, as has been suggested (Baskin-Sommers, Curtin, & Newman, 2011; Newman & Baskin-Sommers, 2011; Newman, Curtin, Bertsch, & Baskin-Sommers, 2010; Wolf et al., 2012), this could enhance overshadowing, thereby reducing learning about the context. For example, Wolf and colleagues report that psychopathic individuals are less susceptible to 'attentional blink', a measure of how detrimental contextual distractors are on responding to a target cue in a visual search paradigm. The reduced attentional blink indicates that contextual distractors interfere less with the performance of the psychopathic participants. The results of the second-order tests in Experiments 3 and 5 also support this suggestion that high psychopathy trait individuals are more resistant to interference by other cues.

Comparison of different contingency conditions facilitates assessment of participants' ability to learn when they are and are not in control of the occurrence of the outcome, and so their ability to monitor the contingency between their actions and their consequences. This was tested by asking participants about their causal control as well as the causal control that is outside their influence and that is attributed to the context, whilst simultaneously providing a measure of response rates and behaviour. In addition, as psychopathic traits have been shown to be predictive of enhanced inhibitory shift learning (Experiments 3-5), it is suggested that psychopathic traits may predict more accurate perceptions of control in the 'shifting' zero contingency condition. Therefore,

comparing participants' sensitivity to the difference between a zero and a positive contingency was also done to provide insight in to the source of the altered shift learning predicted by psychopathic traits in Experiments 3-5.

In addition to testing whether participants could discriminate the two contingencies, I also varied the valence of the outcome to test whether response-outcome contingencies were acquired independent of the affective value of the outcome. As such, participants were exposed to either a neutral or a negative outcome, which was an aversive auditory cue that previous research has shown to elicit strong aversion (Neumann, Waters, & Westbury, 2008; Neumann & Waters, 2006). Outcome valence has previously been shown to have an effect on response rates and reaction times, but not judgements of the causal relation between actions and the outcome (Chatlosh et al., 1985). Given the previous dissociation between the effects of outcome valence on ratings and responses (Chatlosh et al., 1985), I predicted that changes in outcome valence would impact participants' response rates and reaction times, but not their contingency ratings. Varying the valence of the outcome also enabled me to test the predictions of the outcome processing accounts (e.g., R. J. R. Blair, 2005; Lykken, 1995), according to which psychopathic individuals should be unaffected by the outcome valence.

As in the previous experiments, I focused on two of the subscales of the Psychopathic Personality Inventory Revised (PPI-R; Lilienfeld & Widows, 2005): FD and SCI. As dominance captures the exertion of control and influence (Patrick et al., 2009), FD was predicted to be associated with altered perceptions of control, specifically a dichotomy between improved action-consequence perception, and impaired context-consequence perception. Given the effect of behaviour on perceptions of control (Matute, 1996), I tested the direct relationship between FD and perceptions of control, and if this was mediated by responses and/or reaction times. SCI was predicted to relate to

increased behavioural effects particularly in response rates and reaction times. As response behaviour can influence perceptions of control (Matute, 1996), it follows that SCI may indirectly affect perceptions of control via its affect on response rates and reaction times. The third subscale (C) was not expected to predict performance on this type of non-social task.

5.2. Method

5.2.1. Participants

Power analysis conducted on the basis of pilot work comparing action judgements in a zero and a positive contingency ($\mu_1 = 31.17$; $\mu_2 = 7.17$; $\sigma = 35.94$) indicated that 36 participants were needed in each condition to have 80% power for detecting an effect when employing the traditional .05 criterion of statistical significance. Seventy-five participants, including students and non-students, took part in the study and received compensation for their time. Three (female) participants were excluded for failing to follow the task instructions. Seventy-two participants were assigned randomly to the two groups, one using the aversive outcome (24 females, 12 males; Age $M = 20.33$, $SD = 1.39$), and the other the neutral outcome (25 females, 11 males; Age $M = 22.81$, $SD = 5.62$). Ethical approval for the study was granted by the University of Oxford Research Ethics Committee, reference number MSD-IDREC-C2-2013-008.

5.2.2. Design

Participants' overall sensitivity to the contingency manipulation and the outcome valence manipulation was assessed using a mixed factorial design. The between-subject variable was outcome valence (neutral vs. aversive) and the within-subject factor was the contingency (zero vs. positive). The dependent variables were: rates of key presses; reaction times to press the key; judgements of control participants perceived they had

over the occurrence of the outcome (action judgements); and, how much control they perceived the context had (context judgements). Sex and the contingency order counterbalancing (order 1: positive contingency, zero contingency vs. order 2: zero contingency, positive contingency) were included as independent variables in the analysis. As in the previous experiments, zero-order correlations and mediation analyses were also conducted to assess whether FD or SCI predicted contingency sensitivity expressed through either responding (key presses and reaction times) or perceptions of control (action and context judgements).

5.2.3. Materials

Psychopathic Personality Inventory (Revised; PPI-R; Lilienfeld & Widows, 2005). The PPI-R had a good internal consistency (α = .91). The means and standard deviations of PPI-R scores for the two experimental conditions (aversive vs. neutral outcomes) are shown in Table 18.

Table 18

Participant Psychopathy *T*-Scores

	Neutral Outcome Group		**Aversive Outcome Group**	
	N = 36		*N* = 36	
	M	*SD*	*M*	*SD*
Total PPI-R	52.44	8.59	45.94	9.00
Self-Centred Impulsivity	51.53	7.73	47.36	7.79
Fearless Dominance	52.17	9.64	46.19	11.25
Coldheartedness	50.94	9.37	47.81	12.54

Note: The mean *t*-score is 50.

Contingency Learning Task. This task assessed participants' perceptions of whether an outcome was controlled by their actions or by external factors (i.e., the context). Participants were given the choice to perform an action (pressing spacebar). As shown in Figure 22, the probability of an outcome given a response was programmed to be either greater following an action than the absence of an action, or to be the same in the presence or absence of the action. This overall relation is represented by ΔP, a measure of the one-way association between binary events, in this case the action and the outcome (Allan, 1980).

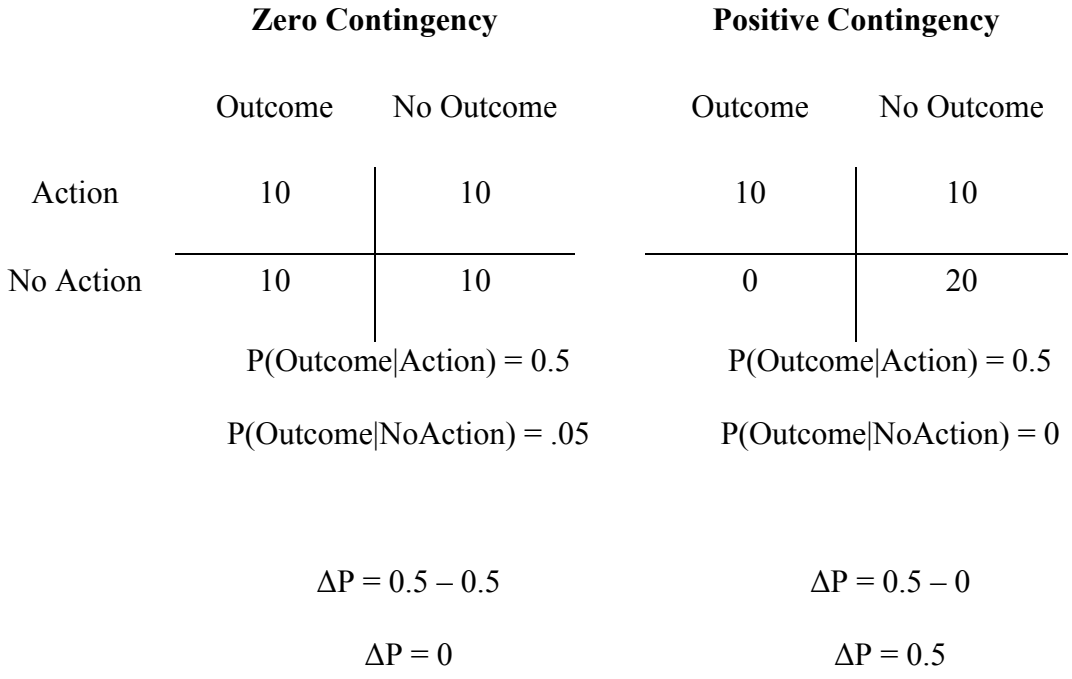


Figure 22. ΔP represents the contingency between an action and an outcome. The figure shows the frequency of trial types in the zero contingency condition (ΔP = 0) and the positive contingency condition (ΔP = .5).

ΔP is calculated as the difference in the probability that the outcome will occur in the presence of the action [p(Outcome|Action)] and the probability of the outcome

occurring in the absence of the action [p(Outcome|NoAction)]. Like a correlation coefficient, the difference between these two probabilities, which varies from -1 to +1, indicates whether the probability that the outcome occurs given an action is made [p(Outcome|Action)] exceeds that of the outcome occurring given no action was made [p(Outcome|NoAction); Allan, 1980].

In the positive contingency condition there was a 50% probability of the outcome occurring when participants responded [pressed the key; p(Outcome|Action) = .5] and a 0% chance of the outcome occurring when they did not [p(Outcome|NoAction = 0], meaning that the positive contingency was .5 ($\Delta P_{positive} = .5 - 0; \Delta P_{positive} = .5$). In the zero contingency condition there was a 50% chance of the outcome occurring regardless of whether the participants responded [p(Outcome|Action) = .5] or not [p(Outcome|NoAction) = .5; $\Delta P_{zero} = .5 - .5$; $\Delta P_{zero} = 0$], meaning that the contingency was ambiguous and constantly reversing.

In order to test whether outcome valence mediated judgement accuracy two different outcomes were used: an auditory stimulus that was developed to be aversive (sounding like a chalk board screech; Neumann et al., 2008; Neumann & Waters, 2006) and a comparison neutral musical auditory stimulus, previously used in this task (e.g., Msetfi et al., 2013). Half of the participants received the aversive stimulus and half received the neutral stimulus for both of their contingency conditions.

Participants received 40 trials of each contingency in a counterbalanced order (.5 first or 0 first). Each trial was 5 seconds long. During the first three seconds of each trial, participants had the opportunity to make a response. Following this, and depending on the programmed probability, the outcome either did or did not occur for two seconds. There was then a three second inter-trial interval before the next trial began. After 20 trials (half-way through the trials for each contingency condition) participants made a

judgement on a sliding scale with +/-1-unit increments from 'totally prevented' (-100) to 'totally controlled' (+100) as to the control their action had over the occurrence of the outcome. Participants also made a judgement using the same scale regarding their perception of the control that the context had over the outcome. Both the action and the context judgements were repeated at the end of the forty-trials for each contingency condition.

The causal cover story suggested that participants were to test for a fault in a music system in a house, with different rooms in the house acting as the two contingency contexts. Using this same cover story, Msetfi and colleagues (Byrom et al., 2015; Msetfi et al., 2005, 2013) have shown that participants are generally accurate regarding their causal control. Participants were shown a 'remote control' on the screen and told that they would have the opportunity to turn on the music in different virtual rooms. Participants were informed that the system was faulty and that sometimes the music might turn on without the button being pressed, so they were encouraged to both press and not press in order to learn their relative effectiveness. Participants that either pressed on all of the trials or on none of the trials were removed from the analysis (n = 3). Participants were assigned to receive either the neutral or the aversive outcome, both of which were two-second auditory clips relevant for the cover story. The order of presentation of the two contingency conditions was counterbalanced and they were distinguishable by distinct virtual contexts (pictures of rooms) relevant for the cover story that implied that the two conditions were independent.

5.2.4. Procedure

Having given their informed consent, participants first completed the Contingency Learning Task followed by the PPI-R (S. O. Lilienfeld & Widows, 2005) and were then debriefed. In the aversive outcome condition, the aversive sound was played to

participants and an opportunity to be excluded from the study was given before commencement. All participants agreed to proceed.

5.2.5. Statistical Analysis

The dependent measures were participants': number of key presses (actions); their reaction times; their action judgements; and, their context judgements. These dependent variables were analysed in relation to the within-subjects variable of contingency condition (zero vs. positive) and the between-subjects variable of outcome valence (neutral vs. aversive) using mixed model analyses of variance to assess overall learning effects. Sex and contingency counterbalancing order were also included as between-subjects factors.

Having established whether overall participants were sensitive to the experimental manipulations of contingency and outcome valence, the relationship between the dependent variables and psychopathy traits was assessed. Zero-order correlations were calculated between each of the dependent variables, and the overall psychopathy score (total PPI-R score), plus each of the PPI-R factors (SCI, FD and C). For this, difference scores between the zero and positive contingency conditions were calculated for each of the variables: for key presses this was done by subtracting the key presses in the zero contingency from those in the positive contingency, meaning that a positive score indicated more key presses in the positive contingency condition than in the zero contingency condition; for reaction times this was done by subtracting the positive contingency RT from the zero contingency RT, meaning that a positive score indicated participants were faster to respond in the positive contingency condition than in the zero contingency condition; for action judgments this was done by subtracting the zero contingency action rating from the positive contingency one, meaning that a positive score indicated a greater action rating in the positive contingency condition than in the

zero contingency condition; and, for context judgments this was done by subtracting the positive contingency context rating from the zero contingency one, meaning that a positive score indicated a greater context rating in the zero contingency condition than in the positive contingency condition.

Multiple mediation analyses with mediators operating in serial were then conducted to test the hypotheses that: FD traits are related to altered perceptions of control (action judgements and context judgements); and, that SCI traits are related to altered behaviour (key presses and reaction times), and so may indirectly affect perceptions of control. As in previous experiments, these relationships were tested using a non-parametric bootstrapping procedure (PROCESS; Preacher & Hayes, 2008) which estimates the indirect effects of multiple mediator variables, and in this case was based on 1000 bootstrap resamples with 95% bias corrected confidence intervals. I tested the direct relationships between SCI/FD and the dependent variables of action and context judgements, plus the indirect relationships between these variables and the possible mediating factors of key presses or reaction times. As behaviour can alter perceptions of control (Matute, 1996) key presses and reaction times were included as mediating variables. Sex, outcome valence condition (neutral vs. aversive) and contingency counterbalancing order were controlled for in the models.

An alpha level of .05 was used throughout, and *t*-tests were only conducted following significant results of ANOVA and were corrected using Bonferroni. As in previous experiments, 95% confidence intervals are reported for ANOVA, *t*-test and mediation analysis where $p < .05$. Also consistent with previous experiments, these confidence intervals are reported to two decimal places, meaning that lower confidence intervals greater than zero, but less than .005 are reported as zero (.00). To distinguish these from confidence intervals equal to or less than zero the latter will be explicitly

flagged, and the implication of a confidence interval that crosses zero on the interpretation of any difference between means will be discussed.

5.3. Results

5.3.1. Overall Learning Effects

Both experimental manipulations were effective: participants were sensitive to the within-subjects experimental manipulation of contingency condition, which was reflected in their reaction times and their action and context judgements (shown in Figure 23); and the effect of the between-subjects manipulation of outcome valence was reflected in participants' key presses. As shown in Figure 24, FD predicted greater sensitivity to the contingency between one's actions and the outcome, but reduced sensitivity to the contingency between the context and the outcome. Psychopathy traits did not predict differences in response rates or speed.

Key Presses (Actions). A significant interaction between contingency condition and outcome valence, $F(1,64) = 11.25$, $p = .001$, $\eta_p^2 = .15$, 95% CI (.03, .31), showed that key presses were higher in the aversive outcome condition, but only in zero contingency condition, (aversive outcome: $M = 26.56$, $SD = 7.52$; neutral outcome: $M = 23.00$, $SD = 3.53$), $t(70) = 2.57$, $p = .012$, $\eta_p^2 = .09$, 95% CI (.01, .22), and not in the positive contingency condition, (aversive outcome: $M = 25.39$, $SD = 7.31$; neutral outcome: $M = 24.58$, $SD = 3.32$), $t(70) = .60$, $p = .549$, $\eta_p^2 = .01$, 95% CI (.00, .08). No other main effects or interactions were significant, all Fs < 2.35, all ps > .13.

Reaction Times (ms). Overall participants responded faster in the positive contingency condition (aversive outcome: $M = 796.86$ms, $SD = 222.98$; neutral outcome: $M = 783.47$ms, $SD = 213.74$) than in the zero contingency condition (aversive outcome: $M = 919.50$ms, $SD = 258.74$; neutral outcome: $M = 854.50$ms, $SD = 296.48$), $F(1,64) =$

6.34, p = .014, η_p^2 = .09, 95% CI (.01, .24). This was unrelated to outcome valence, and no other interactions were significant, all Fs < .95, all ps > .33.

Action Judgements. As shown in the upper panel of Figure 23, participants were sensitive to the contingency manipulation, and judged their actions to be more related to the outcome in the positive contingency condition (aversive outcome: M = 31.75, SD = 25.08; neutral outcome: M = 47.81, SD = 16.68) than in the zero contingency condition (aversive outcome: M = 3.31, SD = 32.12; neutral outcome: M = 8.42, SD = 22.71), $F(1,64) = 62.38$, $p < .001$, $\eta_p^2 = .49$, 95% CI (.31, .61). This effect was not related to outcome valence, and all interactions were not significant, all Fs < .80, all ps > .36.

Context Judgements. Shown in the lower panel of Figure 23, participants' external attributions of control were sensitive to the contingency manipulation, and they attributed significantly greater control to the context in the zero contingency condition (aversive: M = 5.36, SD = 37.81; neutral: M = 19.42, SD = 29.59) than in the positive contingency condition (aversive: M = -29.72, SD = 52.53; neutral: M = -26.47, SD = 51.68), $F(1,64)$ = 25.13, $p < .001$, $\eta_p^2 = .28$, 95% CI (.11, .44). This finding was not affected by outcome valence condition, and no other interactions were significant, all Fs < 2.64, all ps > .11.

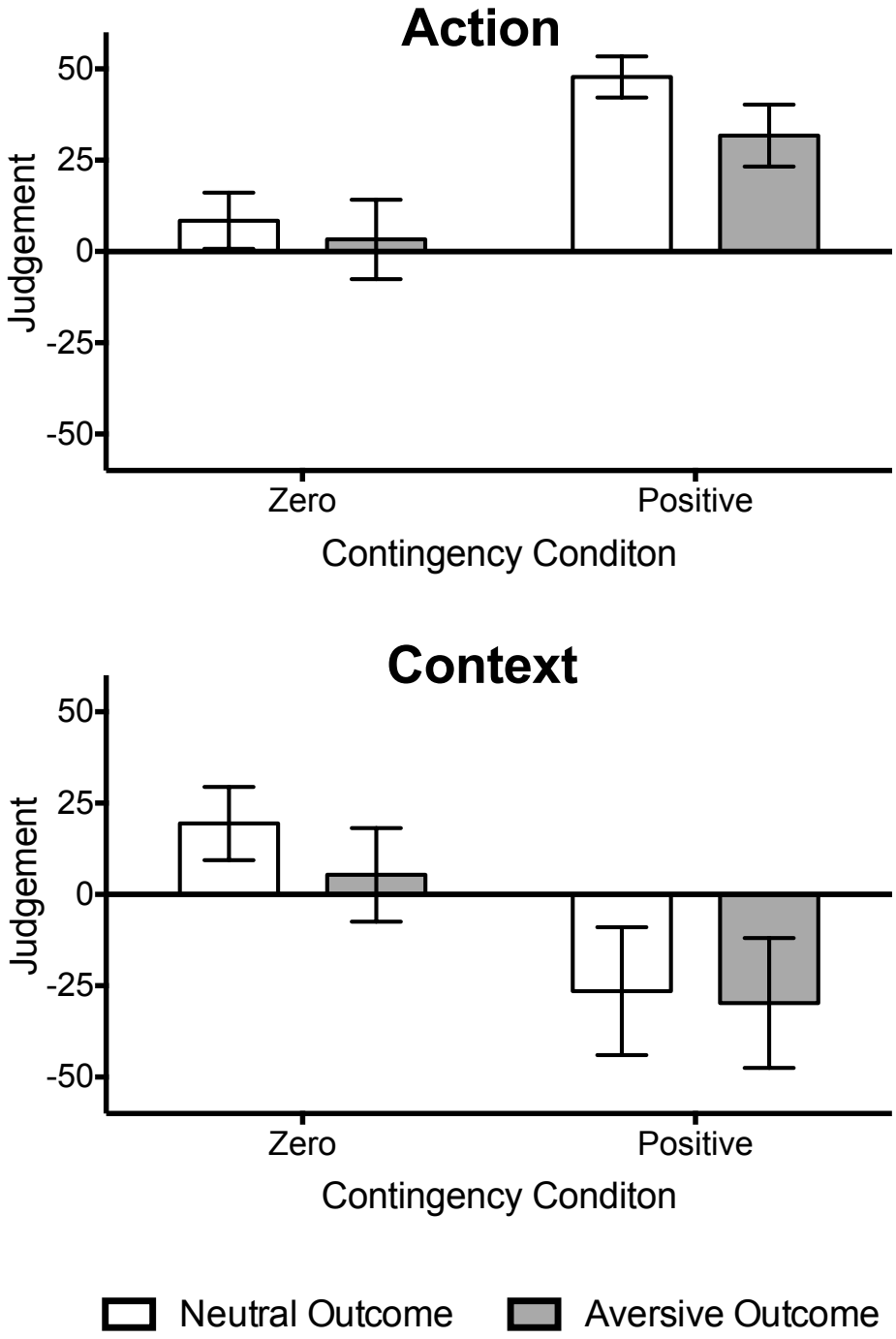


Figure 23. Attribution of control to one's actions was significantly higher in the positive contingency condition than in the zero contingency condition. Conversely, attributions of control to the context were significantly lower in the positive contingency condition than the zero contingency condition. Outcome valence did not significantly affect these results. Errors bars represent 95% confidence intervals.

5.3.2. Relationship Between Psychopathy and Dependent Variables

The zero-order correlations between total psychopathy score, the factors of SCI, FD and C, and the dependent variables are shown in Table 19. As expected, each of the three PPI-R factors was strongly related to the overall PPI-R score. Total score was negatively related to a difference in context judgements between the two contingency conditions (r = -.27, p = .024). FD was positively related to a difference in causal action judgements between the two contingency conditions (r = .27, p = .022), and negatively related to a difference in context judgements (r = -.28, p = .017). SCI was not significantly related to any of the dependent measures.

Table 19

Zero-Order Correlations

	Total PPI-R	**SCI**	**FD**	**C**
Total PPI-R	**-**			
SCI	**.508*****	-		
FD	**.595*****	.114	-	
C	**.324****	-.074	.162	-
KP	.058	-.019	.093	.063
RT	-.163	-.131	-.023	-.013
Action	.066	-.031	**.269***	.044
Context	**-.266***	-.098	**-.280***	-.022

Note: Total PPI-R, SCI, FD and C represent the *t*-scores for each variable. KP = Key Presses (actions); RT = Reaction Time (ms); Action = Action Judgements; Context = Context Judgements.

*** p < .001; ** p < .01; * p < .05.

As shown in Figure 24, participants with higher FD traits showed greater contingency sensitivity for the action-outcome contingency [β = .76, 95% CI (.02, 1.50), $t(70) = 2.04$, $p = .045$, $\eta_p^2 = .06$, 95% CI (.01, .19)], but reduced contingency sensitivity for the context-outcome contingency [β = -2.17, 95% CI (-3.63, -.72), $t(70) = -2.98$, $p = .004$, $\eta_p^2 = .11$, 95% CI (.01, .26)].

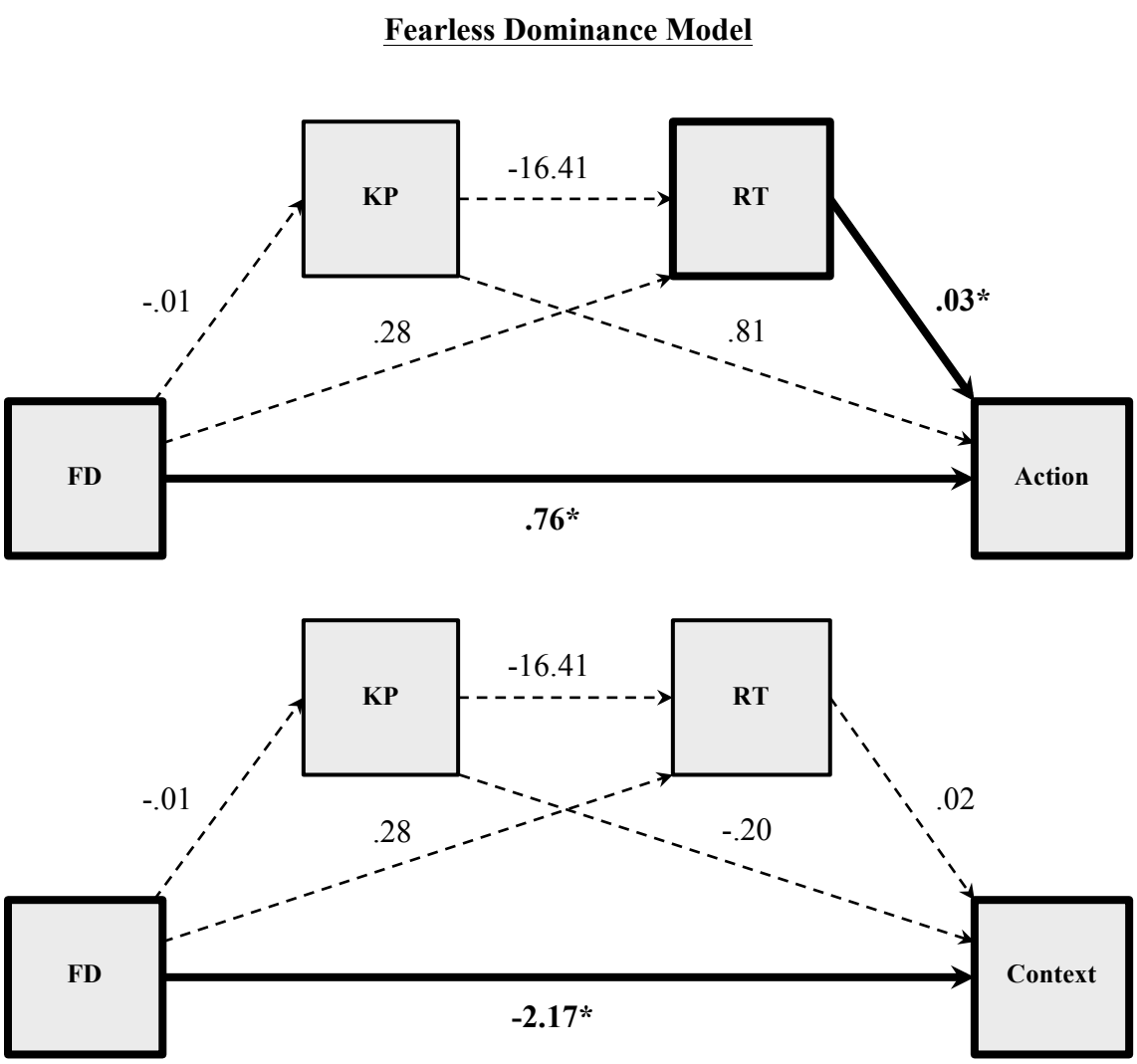


Figure 24. The numbers shown represent β values, signalling the strength of a relationship between two variables. KP = Key Presses; RT = Reaction Time; Action = action judgement difference; Context = context judgement difference. FD predicted greater action-outcome but reduced context-outcome contingency sensitivity.

* $p < .05$

Neither key presses nor reaction times mediated the relationship between FD and action or context judgements, thereby indicating that these altered perceptions of control were independent of response behaviour. Reaction times were significantly positively related to action judgements [β = .03, 95% CI (.00, .06), $t(70) = 2.28$, $p = .026$, $\eta_p^2 = .07$, 95% CI (.01, .20)]. The observed direct relationship between FD and perceptions of control occurred independent of sex, outcome valence and counterbalancing condition.

SCI did not directly or indirectly predict either action judgements (all βs < .25, all 95% CIs (< .01, > .05) or context judgements (all βs < .84, all 95% CIs (< .02, > .14). SCI traits did not alter behavioural control (key presses and reaction times), nor did they affect perceptions of control. Key presses appeared to affect reaction times [β = -17.50, 95% CI (-34.81, -.19), $t(70) = -2.02$, $p = .048$, $\eta_p^2 = .06$, 95% CI (.00, .18)], and reaction times appeared to affect action judgements [β = .03, 95% CI (.00, .06), $t(70) = 2.15$, $p = .035$, $\eta_p^2 = .06$, 95% CI (.00, .19)]. However, both the 95% confidence intervals about the effect sizes crossed zero, suggesting that key presses did not affect reaction times, nor did reaction times affect action judgements.

5.4. Discussion

These results suggest that the FD factor of psychopathy is related to judgements of personal control and may contribute to agentic, self-focused consequence learning. This was characterised by: an increased sensitivity to the consequences of individual actions, and reduced sensitivity to the contingent consequences associated with the context. These findings were unaffected by outcome valence, and were not mediated by behaviour (key presses and reaction times), though overall trends suggested that outcome valence affected response behaviour, indicating that participants were sensitive to the outcome valence manipulation. SCI was not related to altered behaviour, or to altered perceptions

of control either directly or indirectly. Importantly this instrumental learning task does not seem to be simply assessing differences in response rates or reaction time but rather perceptions of causal control that are somewhat independent of these behavioural measures (Matute, 1996).

Consistent with previous research, participants were able to distinguish controllable and uncontrollable outcomes (e.g., Msetfi et al., 2013; Shanks, 2007; Wasserman, Elek, Chatlosh, & Baker, 1993). The two contingency conditions shared the same number of action-outcome pairings and the same probability of the outcome given the action [p(Outcome|Action)]. But, the conditions differed in terms of the total number of outcomes that occur (outcome density), which was higher in the uncorrelated condition (for a discussion see: Vallée-Tourangeau, Murphy, Drew, & Baker, 1998). In many studies, increases in outcome density are associated with illusions of control, whereby a greater number of outcomes leads to the misperception of greater control (e.g., Matute, 1996; Msetfi et al., 2005; Musca, Vadillo, Blanco, & Matute, 2010). In this experiment, illusions of control might have been expected to work against the experimental hypothesis by increasing perceived control in the zero contingency condition, which had twice as many outcomes. Yet contingency sensitivity was still observed, offering strong and conservative evidence that participants were able to distinguish between a controllable and an uncontrollable situation.

This ability to distinguish contingent and non-contingent conditions was positively predicted by FD, suggesting a greater sensitivity to the relationship between one's own actions and the consequences of these actions. However, FD was negatively predictive of the ability to distinguish context-consequence contingencies, meaning that FD predicted a reduced tendency to attribute outcomes to external sources even when the outcomes are aversive. This supports the suggestion that psychopathy is characterised by poor

contextual processing (Hoppenbrouwers et al., 2015), and the hypothesis that psychopathic individuals' learning is more resistant to interference by other cues (e.g., overshadowing; Experiments 3 & 5; Wolf et al., 2012).

Psychopathic traits predicted enhanced learning about one's actions in the 'shifting' zero contingency and the positive contingency, and reduced influence of the context. These results suggest a possible source for the enhanced shift learning predicted by psychopathic traits in Experiments 3-5: namely that the failure to incorporate contextual information into contingency monitoring affects the associability of a cue, leading to greater shift learning. In addition, it was previously discussed that one function of contingency sensitivity may be to allow adaptive interaction with the environment, so as to ensure beneficial consequences for the self (Msetfi et al., 2013). As such, heightened action-consequence contingency sensitivity may also enable high FD individuals to manipulate their environment to suit personal needs, which hints at a possible mechanism for the egocentricity and narcissism associated with FD traits (Benning et al., 2005, 2003; Patrick et al., 2009).

In contrast, the reduced context processing shown here may limit the adaptability of processes dependent on context-specific action-outcome associations, such as those commonly measured in tests of impulsivity (e.g., Budhani et al., 2006; He et al., 2011; Munro et al., 2007b; Newman et al., 1987), which may result in psychopathic individuals appearing to show poor contingency monitoring. In much of the literature the opportunity to perform an action is signalled by a cue (e.g., Brazil, Hunt, et al., 2013; Newman & Kosson, 1986), which can be likened to the contextual cues in the current experiment as both signal unique response-outcome associations. Therefore, reduced context processing could mean that response-outcome associations are not linked to particular stimuli, and so the appropriate association cannot be used to guide behaviour.

The absence of a relationship between SCI and responding in the present sample may be due to participants being advised to respond and inhibit their responses in roughly equal numbers. In other tasks, an opposite approach is taken to encourage the development of a dominant response (e.g., Lapierre et al., 1995; Munro et al., 2007b). As has been reported elsewhere, (e.g., Kiehl et al., 2000), the absence of such response bias leads to more equal allocation of attention, and facilitates more controlled responding (Brazil, Maes, et al., 2013; Kiehl et al., 2000; Sadeh & Verona, 2008). This explanation is consistent with these findings.

It is noted that overall the context judgements in the positive contingency condition were less than zero (M = -28.10), which is not what standard contingency sensitivity models might predict (e.g., Alloy & Abramson, 1979; Mackintosh, 1975; Rescorla & Wagner, 1972). It is possible that this is a result of the rating scale ranging from -100 to +100, thereby encouraging participants to assign a negative value when no control is perceived, though the task instructions did guide against this. A similar pattern of negative context judgements in a positive contingency condition has been observed in other experiments using this task in low outcome density conditions, but not high outcome density conditions (Msetfi et al., 2013). However, Msetfi and colleagues report that the pattern of differences between action and context judgments in zero and positive contingencies was the same across the high and low density conditions, but that the patterns occurred at different points on the rating scale. The positive contingency in the present experiment had fewer outcomes than the zero contingency, and so it is possible that this is why context judgements in the positive contingency conditions are below zero. Future studies might explore this possibility, and should endeavour to ensure that the outcome density is as similar as possible when comparing across contingency conditions.

5.5. Conclusion

The contingency task used here offers a novel way of assessing contingency sensitivity and the balance between attributions of control to the self and attributions of control to the context. FD predicted self-focused consequence learning, demonstrated through increased sensitivity to action-consequence correlation perceptions, but insensitivity to context-consequence correlation perceptions. This imbalance suggests a possible source for the enhanced shift learning seen in relation to psychopathic traits in Experiments 3-5. The results of Experiment 7 are consistent with those of the previous experiments (1-6), and provide additional support for the suggestion that learning in psychopathy is altered due to a narrow processing bottleneck, which reduces the extent to which psychopathic individuals' learning is influenced by other cues.

CHAPTER 6

General Discussion

The aim of this thesis was to test whether psychopathic traits predicted differences in learning, and to determine the conditions under which these differences occurred. The results of Experiments 1-5 showed that it is possible to train associative learning using a go/nogo paradigm. Overall: psychopathic traits were not predictive of differences in response errors to either excitatory or inhibitory cues (except for more accurate responding in Experiment 3); nor were psychopathic traits predictive of differences in learning about excitatory cues as measured by ratings in either direct (shift) or indirect (e.g., super-learning or blocking) tests. However, psychopathic traits consistently predicted greater shift learning to the previously established inhibitor when the contingency change was relatively small, and the cue was weakly predictive. When the contingency change was larger and the cue was more predictive, psychopathic traits did not predict differences in shift learning.

These findings are inconsistent with the predictions of theories of learning and attention, which suggest that shift learning about an inhibitor should be retarded, particularly when the outcome becomes less predictive (e.g., Le Pelley et al., 2011; Mackintosh, 1975; c.f., Pearce & Hall, 1980). However, the results are compatible with the broad principles of theories of psychopathy (e.g., Hamilton et al., 2015; Newman & Baskin-Sommers, 2011), which suggest that psychopathic individuals may have an information bottleneck that reduces their processing capacity, particularly with regards to

peripheral or contextual stimuli. In Chapter 3, I argued that such a bottleneck might account for the differences seen in inhibitory learning due to the increased processing demands required of forming an inhibitory association (e.g., AX-, X+) as compared to an excitatory association (e.g., BV+). I test this in Experiment 6, where participants performed an articulatory suppression task concurrently with the go/nogo task to increase cognitive load demands (Baddeley et al., 1984; Tullett & Inzlicht, 2010). Under these conditions, psychopathic traits were not predictive of learning differences. As load was not a within-experiment manipulation, the conclusions that can be drawn are limited. But the overall main effect of shift learning and the absence of second-order effects is a pattern that more closely mirrors the high psychopathy trait individuals' performance in Experiments 3-5. As such, this appears to suggest that when placed under high cognitive load conditions, the low psychopathy individuals act more like the high psychopathy individuals, though further research is needed to verifying this.

Together the findings of Experiments 1-6 are suggestive of narrow processing, and so reduced contextual processing in psychopathy. This was shown to affect shift learning, and rapid switching of contingencies, an ability that was positively predicted by psychopathic traits. As such, in Experiment 7 I used an instrumental learning task to measure participants' contingency sensitivity to a continuously shifting (zero) contingency and a positive contingency. Participants' judgements of their own control over causing an outcome were contrasted with the control they attributed to the context in order to test the extent to which the context was processed and learnt about. The results showed that psychopathic traits predict greater sensitivity to the contingency between one's actions and the outcome, but reduced sensitivity to the contingency between the context and the outcome. This disparity suggests a possible source for the enhanced shift learning seen in Experiments 3-5, as it has previously been suggested that associations

are evaluated in contrast to other contextual associations (Byrom et al., 2015). Therefore, reduced contextual processing will affect learning about other, more central stimuli. These differences in contextual and information processing appear to be a core component of psychopathy as these studies have shown these effects to be correlated with overall psychopathic traits, and predicted by the factors of SCI and FD. As such, I argue that this narrow processing bottleneck and altered contextual information processing are central features of psychopathy and may underlie psychopathic learning differences.

6.1. Impaired Integration Theory

As outlined in Chapter 1, theories of psychopathy have typically cited either outcome processing problems, resulting in a failure to use affective outcome information to guide goal-directed responding (R. J. R. Blair, 2005; Moul et al., 2012); or, attentional problems resulting in overly focused attention and the failure to process peripheral information (Gorenstein & Newman, 1980; Newman & Baskin-Sommers, 2011). Though both groups of theories have their strengths, neither is able to account for the range and specificity of the differences seen in relation to psychopathic traits: for example, outcome processing accounts cannot explain the context specificity of psychopathic differences (Newman & Schmitt, 1998); and, attentional accounts cannot explain affective processing deficits when the emotional content is central to the focus of attention (R. J. R. Blair et al., 2004).

Hamilton and colleagues (2015) proposed the Impaired Integration Theory of psychopathy in order to amalgamate attentional and outcome processing accounts, and to include a neurobiological perspective. The full principles are outlined in Chapter 1 (1.4.), so only a brief summary will be given here. The Impaired Integration Theory suggests

that psychopathic individuals have shallow information processing, which results in a perceptual bottleneck and a relative failure to process peripheral information. Given the shallow information processing, the theory predicts that the acquisition of associations will occur under conditions akin to those of high perceptual load in controls. This assertion is supported by the results of Experiment 6, which suggest that the addition of a cognitive load task may induce low psychopathic trait individuals to act more like high psychopathic trait individuals. This is also consistent with the poor contextual processing shown in Experiment 7 through the limited formation of associations between the context and the outcome.

However, the Impaired Integration Theory (Hamilton et al., 2015) also predicts that as a result of this shallow information processing, high psychopathy trait participants will be unable to shift their learning about cues, as associations are not easily updated. This is predicted to result in the characteristic psychopathic perseveration, with participants failing to show significant shift learning to excitatory and inhibitory cues. This prediction contradicts the findings of Experiments 1-5 in which psychopathic traits predicted comparable or enhanced shift learning to the inhibitors, and did not predict differences in shift learning to the excitors.

6.2. Theory of Attention

An alternative way of interpreting the present findings is through a theory of attention that refers to cue predictiveness (e.g., Mackintosh 1975). As outlined in Chapter 1 (1.5.), in addition to an error prediction term that captures the disparity between the actual (λ) and the expected outcome (ΣV; e.g., Rescorla & Wagner, 1972), learning is also dependent on how uniquely reliable a cue is at predicting the outcome (Mackintosh,

1975). This associability, denoted as alpha (α), determines how easily the cue is learnt about.

As discussed in Chapter 1, Mackintosh's (1975) theory predicts that shift learning, such as that tested in Experiments 1-6, should initially be facilitated by both a large error prediction term and the high associability of the cue, which is established in the acquisition phase due to it being a good predictor. However, in the shift phase though the error prediction term would still drive learning, the change in the contingency reduces the predictiveness of the cue, which reduces its associability, and so limits subsequent shift learning. Therefore, some shift learning will be seen, but this will not reflect the full extent of the contingency change. This prediction is consistent with the performance of the low psychopathic trait individuals, but not the high psychopathic trait individuals.

Mackintosh's (1975) theory is also able to explain the enhanced shift learning and absence of second-order learning effects that are seen in all participants in Experiment 6. The process of establishing the associability of a cue relies on integrating information in order to determine the cue's relative validity as a predictor of the outcome. If participants' processing is occurring under high load conditions, then they may not have enough capacity to compare and integrate the predictiveness of the other cues present, meaning that they cannot update the cue's associability. This will result in the cue retaining its high associability throughout the shift phase, meaning that more extensive shift learning will be seen as compared to when the necessary processing does occur under low load conditions.

It is possible that this explanation of participants' performance in Experiment 6 is more broadly representative of learning in high psychopathy traits individuals. Interpreting the results of Experiment 6 through Mackintosh's (1975) theory provides a mechanistic account that is broadly compatible with the performance of the high

psychopathy trait individuals in Experiments 1-5, as well as the broad processes described in the Impaired Integration Theory (Hamilton et al., 2015). I suggest that psychopathic traits are predictive of a disruption to both the process of determining a cue's associability, and using this associability to guide learning and behaviour. The ability to explain the specificity of the observed differences will now be discussed in relation to a novel account of psychopathy that incorporates these principles.

6.3. Novel Proposal: The Psychopathy Attention Theory

I argue that the broad features of the Impaired Integration Theory (Hamilton et al., 2015) are compatible with Mackintosh's (1975) model, and that, by integrating the two, I can outline the mechanism that might reflect the learning differences seen in psychopathy. There is considerable evidence that shallow information processing occurs in psychopaths for both affective (Baskin-Sommers et al., 2013; K. S. Blair, Morton, Leonard, & Blair, 2006; Lykken, 1995) and nonaffective stimuli (Hiatt et al., 2004; Wolf et al., 2012). This shallow information processing, likened to a state of high perceptual load (Hamilton et al., 2015), is suggested to result in poor contextual processing (Hoppenbrouwers et al., 2015; Aisbitt, Msetfi & Murphy, submitted). This demonstrates psychopaths' failure to use information outside of the current goal-pursuit to guide learning, which results in 'Impaired Integration' of information.

According to Mackintosh's (1975) theory, cue competition governs learning and a cue's associability is determined by its relative validity as a predictor of the outcome. This process of determining a cue's relative validity is dependent on integrating information about the predictiveness of all the other cues associated with the outcome. Therefore, if psychopathy is characterised by a difference in integrating information, this process of determining a cue's relative validity will be altered. As such, in psychopathic

individuals the value of α may not adjust to reflect the cue's relative validity with the same degree of sensitivity that is seen in non-psychopathic individuals. Specifically, I suggest that psychopathic traits adversely affect individuals' ability to compare predictiveness of a cue ($\lambda - V_A$) to the predictiveness of all the other cues ($\lambda - V_X$), a necessary step in marshalling changes in α. This is suggested to occur given the reduced processing capacity associated with psychopathic traits (Hamilton et al., 2015), and so I argue is particularly related to the process of assessing the predictiveness of the other cues. I suggest that determining the predictiveness of the central test cue is relatively spared, but that psychopathic individuals struggle to compare this to the predictiveness of other cues. This is consistent with the overly goal-focused, central processing tendencies seen in psychopathic individuals at the expense of more peripheral or contextual stimuli (Baskin-Sommers et al., 2011; Wolf et al., 2012).

The specificity of the difference in psychopathic individuals' learning, namely altered shift patterns to inhibitory cues but not excitatory cues can be accounted for using the proposed Psychopathy Attention Theory. As has been suggested (Baskin-Sommers et al., 2013; Hamilton et al., 2015), psychopathic traits appear to relate to reduced processing capacity, which I argue interferes with the practice of determining a cue's relative validity. The process necessary to train an associative inhibitor has been outlined in Chapter 1 (1.7.), and the differences in this process compared to that necessary for training an excitor have been discussed. Particularly relevant is the necessity to combine information across stimuli in order to form an inhibitory association (e.g., AX-, X+), but not an excitatory association (e.g., BV+). Therefore, whilst community psychopathic individuals may be able to cope with this initial processing demand, this will substantially increase (perhaps even double) the processing requirements when determining the relative validity of an inhibitory cue, which may take the processing

capacity beyond its maximum limit. This would result in a specific difficulty in updating the associability of inhibitory cues given the more complex training requirements.

I suggest that this ability is not binary, and instead it appears that psychopathic traits are inversely related to an individual's ability to process and integrate information. This is consistent with Experiments 3-5 showing a linear relationship between psychopathy and altered learning dependent on determining a cue's associability. This linear relationship supports the study of psychopathy in community samples, and is consistent with previous research (e.g., Mahmut et al., 2008) by suggesting that learning and behaviour differences seen in forensic samples are, at least in part, due to a shared mechanism. If, as has been suggested, psychopathic traits are associated with limited processing capacity then there will come a point when this processing capacity restricts their ability to develop an inhibitory association, which depends on integrating information across cues (e.g., AX-, X+). A suggestion of this was seen in Experiment 3, where psychopathic traits predicted a reduced difference in the initial acquisition of an inhibitory association, as compared to a cue simply paired with the absence of the outcome. Though this was not seen in subsequent studies, the effect provides a possible mechanism by which more extreme inhibitory control problems may be explained, for instance differences in the initial acquisition of inhibitory associations that are sometimes seen in forensic samples (R. J. R. Blair et al., 2004; Newman & Kosson, 1986).

I argue that this impaired ability to update a cue's associability is distinct from psychopaths' preserved use of error prediction terms. There is substantial evidence of psychopathic individuals' ability to learn about the relationships between stimuli, responses and outcomes (Brazil, Maes, et al., 2013; Budhani et al., 2006; Kiehl et al., 2000; Verona, Sprague, & Sadeh, 2012), findings consistent with the experiments

reported here. As such, it is broadly accepted that psychopathic individuals can use an error prediction term to form basic associations (Moul et al., 2012).

Evidence to the contrary typically involves the use of affective or emotionally-salient outcomes (e.g., K. S. Blair et al., 2006; Flor, Birbaumer, Hermann, Ziegler, & Patrick, 2002). For instance, Blair and colleagues used a task in which participants were required to choose between two objects associated with different levels of reward or punishment (+/- 100, 200, 400, 800, 1600). The control group showed more rapid learning to cues associated with a higher positive or negative value, whilst the psychopathy group learnt the association between the cues and their respective outcomes at similar rates. This was seen as insensitivity to the value of the outcome, and a failure to use an error prediction term to learn. However, rather than suggesting that it is the processing of affective information that is impaired in psychopathy, it is more parsimonious to consider that the mechanism for changes in associability is altered.

Previous research has shown that the valence of an outcome can influence attention that is directed to cues associated with it, with participants paying more attention to cues that predict a high-value outcome as compared to cues that predict a low-value outcome (Le Pelley et al., 2013). This increased associability of cues that are predictive of a high-value outcome will be reflected through the cues' α. Therefore, if the changeability of the α parameter is disrupted, a high-value cue (either reward or punishment) will not be able to increase the ease with which psychopathic individuals learn about the cue. Though there will be some difference in the value of the error prediction term, the reduced difference in α would result in cues associated with high-value outcomes and those associated with low-value outcomes being learnt about at similar rates. Specifically, I suggest that the error prediction term does not receive influence from changes in associability, and so the affective content of the cues is not able to facilitate learning.

This may provide a more general explanation for learning and attention deficits than have hitherto been suggested.

A similar explanation can account for the poor fear conditioning that is often seen in psychopathy. It is consistently reported that psychopathic individuals show a preserved ability to form associations between stimuli, so-called 'cognitive-information processing', but an impaired ability to acquire conditioned physiological fear (Birbaumer & Veit, 2005; Flor et al., 2002; López et al., 2013). For example, López and colleagues report that students with high levels of FD showed reduced changes in skin conductance, a measure of fear conditioning to an aversive stimulus, but that psychopathic traits were not related to differences in contingency-awareness ratings. The former 'emotional associations' are argued to be reliant on changes in α to facilitate associations beyond those signalled by the error prediction term (López et al., 2013; Mackintosh, 1975). According to Mackintosh's (1975) Theory of Attention, emotionally salient stimuli will have a higher alpha, meaning that the emotional valence of a cue is able to facilitate increased learning. As such, learning about fearful stimuli should be enhanced, as is shown in non-psychopathic individuals. The relative failure of individuals with high psychopathic traits to show conditioned fear responses is, therefore, suggestive of their relative failure to adjust the associability of a stimulus, but their preserved ability to form associations between stimuli, when these are only dependent on the error prediction term.

I propose that these differences can be modelled by considering a trait factor modification to the modified Mackintosh model (Le Pelley, 2004; Mackintosh, 1975) proposed to govern learning (Aisbitt & Murphy, submitted a). I suggest that factor *P* correlates with participants' level of psychopathic traits ($1<P<2$) and propose that the following equations can be used to estimate the adjustment of α in relation to psychopathic traits:

$$\Delta\alpha_A > 0 \text{ if } |\lambda - V_A| < P|\lambda - V_X|$$

$$\Delta\alpha_A < 0 \text{ if } |\lambda - V_A| \geq P|\lambda - V_X|$$

whereby as psychopathic traits increase, P increases, and so consequently reduces the perceived predictiveness of other stimuli when determining the unique predictiveness of a given stimulus. This underestimation of the predictiveness of other cues relative to a given stimulus results in stimuli being perceived as more predictive than they necessarily are, and correspondingly α values that are too high.

It is also possible that psychopathic traits affect the process of determining the predictiveness of the test stimulus itself, as opposed to the predictiveness of all other cues, as the most important point is for there to be an imbalance between these two parts of the equation for determining alpha. For this explanation, the trait factor *P* would need to be inverted, such that higher psychopathic traits led to a smaller *P*, which would increase the likelihood of a positive change in alpha occurring. However, it is not clear why psychopathic traits would affect the determination of the predictiveness of a single, central cue, and not that of a collection of more peripheral, or contextual cues. It is more consistent with the psychopathy literature, in which psychopathic traits are associated with overly focused attention and goal-directed processing at the expense of peripheral stimuli (Baskin-Sommers et al., 2011; Wolf et al., 2012), to suggest that it is the latter aspect of the equation which is affected. Indeed, it is also possible that the process of determining alpha is unimpaired, but that the influence alpha has on learning is disrupted, though again it is not clear how this fits with the current evidence, or the principles of psychopathy theories suggested elsewhere. It is, however, noted that these possible means by which alpha may be disrupted were not explicitly tested here, and instead are derived *post hoc* drawing on the literature and the experiments contained in this thesis.

Either way, the suggestion that psychopathic traits disrupt alpha is consistent with the evidence suggesting that psychopathy involves disruption in specific areas of the brain (e.g., basolateral amygdala), which are involved in allocating attention to stimuli on the basis of their relative predictiveness, and in modulating stimulus processing (see Moul et al., 2012). As previously mentioned, as well as capturing the attentional components, and appearing consistent with neurological explanations of psychopathy, the parameter α is able to explain, at least in part, the emotional-affective component of psychopathy through reduced facilitation of learning about emotionally-salient, affective stimuli. Therefore, the proposed Psychopathy Attention Theory has been developed to incorporate the principles of the Impaired Integration Theory (Hamilton et al., 2015) into Mackintosh's (1975) Theory of Attention and so provide a mechanistic account of learning in psychopathy. In so doing, I have shown that it is possible to account for the results of the series of experiments presented here but also for the inconsistencies in learning and reversal that are reported in psychopathic individuals through the influence of methodological feature on a cue's associability (see also section 1.6.).

6.4. Future Experiments to Test the Psychopathy Attention Theory

One way for future research to test this proposed mechanism is to contrast performance that requires processing of peripheral and central stimuli, and test how predictiveness affects this. If the selectivity of attention is increased by psychopathic traits, then participants with high psychopathic traits should show a greater bias in their pattern of learning as a result of attentional manipulations. One of the simplest phenomena of human associative learning thought to implicate attention is stimulus over-selectivity (Dube & McIlvane, 1999; Dube et al., 1999) or overshadowing. As previously outlined, this occurs when one cue of a stimulus compound comes to exert greater

control over behaviour than other, equally valid, members of that same compound. The attentional account of over-selectivity (Dube et al, 1999) argues that people do not attend equally to all components of the compound during training. Thus, unattended stimuli are not learned about, and so do not come to control behaviour. The critical finding is that, although two stimuli may be equally predictive, one is selected as more important than the other, and so receives more attentional allocation.

If psychopathic traits heighten attentional biases and narrow processing, then high psychopathic traits should result in more pronounced over-selectivity: if the stimuli in a compound are equally predictive, then participants will be more likely to pick one element of each stimulus compound at the start of training and concentrate their attention (and hence learning) on that stimulus across training trials. Conversely, if a compound comprises a predictive and a non-predictive stimulus, psychopathic traits might be expected to be associated with the reduced discrimination of the two stimuli, for instance as was suggested in Experiments 3-5, resulting in more balanced attentional allocation. Future research might wish to investigate this possibility using a procedure similar to one previously used to demonstrate attentional biases in negative mood (e.g., Le Pelley et al., 2011). The extent of over-selectivity could be measured by the absolute difference in ratings received by two stimuli that belong to the same compound. Strong over-selectivity would be indicated if participants were more confident that one of the two elements predicts the outcome.

Learning in these experiments was measured by motoric response and explicit judgement, with most of the effects demonstrated in the explicit measurement. It would be interesting to see whether some of the predictiveness findings were reflected in more covert measures of attention and learning. Eye-tracking, for example, could also be used to monitor overt attention during training, providing another measure of over-selectivity

in terms of the difference in eye-gaze to the more-attended and less-attended cue of each compound. Eye-tracking is a useful measure of attention because attention and eye-gaze are generally tightly coupled (Deubel & Schneider, 1996). For instance, a recent study has shown eye-gaze preference for fearful stimuli, and has even suggested that amygdala activation during this experiment was not caused by the fearful stimuli, per se, but by the amygdala's role in the reflexive attentional shift to the fearful stimuli (Gamer & Büchel, 2009).

Given previous suggestions of altered amygdala functioning in psychopathy (R. J. R. Blair, 2007; Budhani, Marsh, Pine, & Blair, 2007; Gao et al., 2010; Moul et al., 2012) it follows that processes such as reflexive attentional shifts to emotionally salient stimuli may be disrupted. This suggests a possible link between neurobiological, attentional and outcome processing accounts of psychopathy. Following the principles of the Psychopathy Attention Theory outlined here, and taking in to consideration the broader literature on psychopathy, I might predict that participants with higher psychopathic traits would show greater over-selectivity in both overt attention (eye gaze) and their learning ratings, but that this difference would not be seen if the stimuli were emotionally valent, or if the cues were not equally predictive.

6.5. Conceptualisation of Psychopathy

Taken together, the enhanced shift learning, reduced influence of other cues on learning, and altered perceptions of control are consistent with the conceptualisation of FD as a combination of social adjustment skills and so-called 'successful psychopathic' traits, as well as more maladaptive ones such as narcissism and disregard for people and things external to the self (Benning et al., 2005; S. O. Lilienfeld & Widows, 2005; Patrick et al., 2009). As discussed in Chapter 1, the literature is divided over the role of

positive adjustment features (captured by FD) in the conceptualisation of psychopathy (e.g., Miller & Lynam, 2012; Patrick et al., 2009). One opinion is that psychopathy is only characterised by negative or maladaptive traits and behaviours (Hare & Neumann, 2008). Proponents of this view argue that psychiatric disorders are typically defined in terms of maladjustment (Livesley, 2007), and so the conceptualisation of psychopathy should follow the same pattern (Hare & Neumann, 2008).

However, an opposing view (Patrick et al., 2009; Patrick, 2006) is that the negative conceptualisation of psychopathy has moved away from Cleckley's (1982) original depiction of psychopathic personality, which included four positive adjustment features: superficial charm; absence of delusions and irrational thinking; suicide immunity; and, the absence of nervousness and anxiety. One of the most widely used tools for assessing psychopathy, the Psychopathic Checklist (Revised; Hare, 1991), was based on Cleckley's conceptualisation of psychopathy, yet does not include any of the four positive adjustment features he described. Indeed, the PCL-R has been criticised for blurring the line between the construct of psychopathy and how this construct is assessed by becoming too directive in the factors it includes (Cooke & Michie, 2001; Patrick, 2006). Furthermore, the failure to include positive adjustment features of psychopathy has led to some *post hoc* attempts to incorporate these principles in order to explain differences seen within psychopathy samples (e.g., Newman et al., 1990; Newman & Schmitt, 1998).

Take, for instance, the absence of nervousness or anxiety, a construct that is not represented in the PCL-R. Newman and Schmitt (1998) used a go/nogo task to test forensic participants' impulsivity. Participants were allocated to groups on the basis of their psychopathy score (PCL-R), and, importantly, on their levels of anxiety, thereby creating four experimental groups based on: high or low psychopathy with high or low anxiety. The argument for this was that although psychopathy and anxiety are

'independent' they are associated with the same psychological concepts, such as punishment insensitivity, and that conceptualisations of psychopathy suggest it should not involve high anxiety levels. As the interaction between impulsivity and anxiety can alter responding (Corr, 2002), group allocations on the basis of both psychopathy and anxiety affect results. For example, in Newman and Schmitt's study only the low anxiety psychopaths were shown to be more impulsive than the low anxiety controls, but their high anxiety counterparts did not significantly differ, and there were no differences when the groups are collapsed across anxiety.

This example shows how differences in, for example, impulsivity may relate only to specific sub-types of psychopathy and not to others. By failing to include the positive adjustment features, the PCL-R (Hare, 1991) was not able to distinguish these sub-types without the additional information gained through an anxiety assessment. In contrast, a measure that does include positive adjustment features, such as the Psychopathic Personality Inventory (Revised; PPI-R; Lilienfeld & Widows, 2005) may have captured these unique group profiles without the need for an additional measure. For instance, whilst both the FD and SCI factors of the PPI-R are associated with behavioural and social maladjustment, FD is uniquely predictive of reduced fearlessness and anxiety. Therefore, were a measure such as the PPI-R, which includes positive adjustment features, used then the differences between the high/low psychopathy and high/low anxiety groups might have been captured by differences between individuals scoring high on either FD or SCI.

Furthermore, by allocating participants to experimental groups on the basis of other criteria (e.g., anxiety), this fails to consider the influence these other criteria may have on processes that are altered in psychopathy. For instance, anxiety levels may affect: attentional allocation (Eysenck, 1988; Fox, 1993), particularly when there is an affective

value to cues (Waters, Nitz, Craske, & Johnson, 2007); and, sensitivity to outcome information (Torrubia, Ávila, Moltó, & Caseras, 2001). It has been suggested that both attention and outcome processing are disrupted in psychopathy (R. J. R. Blair, 2005; Gorenstein & Newman, 1980; Hamilton et al., 2015; Moul et al., 2012). Therefore, by allocating groups on the basis of psychopathy and separate criteria, such as anxiety (e.g., Newman & Schmitt, 1998), as opposed to using a measure which captures this and its relation to psychopathy in one, it may influence the results of studies looking to address the attentional and outcome processing capacity of psychopathic individuals.

As such, I argue that it is ultimately the imbalance between processes and abilities that is defining, and that these imbalances may occur as a result of over-developed abilities (e.g., increased sensitivity to action-consequence contingencies), under-developed abilities (e.g., reduced sensitivity to context-consequence contingencies), or, as was shown in the experiments presented here, a combination of both. Therefore, I suggest that it is important to include both adaptive and non-adaptive features in the conceptualisation of psychopathy so as to comprehensively assess the imbalance in processes that underlie observable differences.

6.6. Assessing Inhibition

The novel go/nogo task developed in Experiments 1-6 has implications on psychopathologies other than psychopathy. Indeed, a number of psychopathologies are attributed (at least in part) to a failure of inhibition (Nigg, 2000). For example: substance abuse (Nigg et al., 2006); eating disorders (Houben, 2011); attention-deficit/hyperactivity disorder (Barkley, 1997; Oades et al., 2008; Winstanley, Eagle, & Robbins, 2006); bipolar disorder (Moeller, Barratt, Dougherty, Schmitz, & Swann, 2001; Swann, Anderson, Dougherty, & Moeller, 2001; Swann, Pazzaglia, Nicholls, Dougherty, &

Moeller, 2003); and, alcoholism (Moeller et al., 2001; Nigg et al., 2006) are all considered to be characterised by altered inhibitory control. Weak self-regulation is common to these psychopathologies, restricting the capacity for on-going regulation of behaviour. For example, substance abuse is associated with impulsive choices, whereby an individual will choose a small, immediate reward over a large reward that occurs later (Perry & Carroll, 2008).

Typical experimental procedures for studying inhibition in such psychopathologies employ either predictive causal learning tasks (e.g., Gillan et al., 2014), or active go/nogo or stop-signal tasks (e.g., Brazil, Maes, et al., 2013; Kaufman, Ross, Stein, & Garavan, 2003). For instance, there is evidence that on a go/nogo task: children with ADHD (Iaboni, Douglas, & Baker, 1995); cocaine users (Kaufman et al., 2003); alcoholics (Kamarajan et al., 2005); and, psychopathic individuals (Brazil, Maes, et al., 2013) all showed increased response errors, suggestive of poor inhibitory control. However, it has also been shown that reversal learning is altered in psychopathic individuals (Brazil, Maes, et al., 2013; He et al., 2011), but that reversal learning is unrelated to alcohol misuse (Gullo, Jackson, & Dawe, 2010). Therefore, whilst a number of psychopathologies involve inhibitory deficits such as those measured by a basic go/nogo task, the nature of these may vary, and it appears that more complex designs are needed to tease apart these differences.

The novel go/nogo task developed here provides a means by which to assess the specificity of these inhibitory differences by testing the impact that excitatory and inhibitory learning has on other cues, as well as on subsequent learning about the original cue. The results of the present experiments indicate that complex associative designs may be incorporated into a go/nogo procedure in order to facilitate the more extensive exploration of inhibitory and excitatory processes. Therefore, this novel go/nogo task can

be used to explore the precise nature of inhibitory differences that characterise distinct psychopathologies broadly regarded as 'inhibitory control disorders'.

6.7. Negative Super-Learning

In Experiments 1-6 participants' ratings reflected an ability to form excitatory and inhibitory associations expressed in ratings of causal effectiveness. This builds on previous research by suggesting that a go/nogo procedure can be used to train associations, something that was previously suggested to be somehow unique to stop-signal tasks (Bowditch, Verbruggen, & McLaren, 2015). These experiments also examined second-order learning effects, whereby learning about a stimulus influences subsequent learning about another stimulus (e.g., blocking, Kamin, 1969; super-learning, Rescorla, 1971b).

Across the experiments I show that participants distinguish between: novel cues trained with a cue predictive of the same outcome; and, novel cues trained with a cue predictive of a different outcome. For instance, a novel cue reinforced in the presence of a previously trained inhibitor acquired stronger excitatory strength than other excitatory cues. This is suggestive of positive super-learning, thereby supporting previous demonstrations of positive super-learning in humans (e.g., Lovibond et al., 1988) and showing that this second-order effect can be induced in a go/nogo paradigm. Furthermore, and to the best of my knowledge, this is the first evidence of so-called 'negative super-learning', whereby a previously trained excitor can facilitate inhibitory learning about a novel cue when the two are subsequently trained in compound as inhibitory (Aisbitt & Murphy, submitted b). This finding is potentially important to the

theoretical conceptualisation of inhibition as it indicates symmetry in super-learning (e.g., Rescorla & Wagner, 1972).

However, given the limitations in the comparison cues that were discussed throughout Chapter 3, future research must further examine the extent to which this positive and negative super-learning can be found using different comparison cues. For instance, a design might compare super-learning in the presence of the potentially overshadowing cue (e.g., AY+, A-) and the absence of the potentially overshadowing cue (AY+), with appropriate comparison cues, DS+, D+ and DS+ respectively. Despite these limitations, and the need for further research, Experiments 1-5 suggest that participants can form enhanced excitatory and inhibitory associations independent of simple outcome pairing, and so provide preliminary evidence of the positive and negative enhancement and interference effects (i.e., super-learning or blocking), though the extent to which this can decisively be interpreted as super-learning remains to be seen.

If future research can still demonstrate the 'super-learning' effects reported here when using alternative, and more stringent, comparison cues, then these findings have potential implications for inhibitory training programs. Modified go/nogo and stop-signal tasks have been used to treat impulsivity and poor inhibition, for example, in eating disorders (Houben, 2011; Veling, Aarts, & Stroebe, 2013) and alcoholism (Bowley et al., 2013; Jones & Field, 2013). In one study on binge eating, participants were found to be significantly less likely to select a "palatable unhealthy food" if its image had been previously paired with stop-stimuli, signalling the need to withhold responding (Veling et al., 2013). Invariably, these training programs are designed to improve inhibition in individuals who show weak, or attenuated inhibition, for example in their ability to resist drinking (too much) alcohol (Bowley et al., 2013; Jones & Field, 2013). As such, the demonstration that adopting a super-learning procedure can enhance the same inhibitory

induction used in these training programs could have a significant impact on their effectiveness.

Super-learning effects would also work for inhibitory control disorders characterised by heightened levels of inhibition. For example, within the eating disorders category, bulimia nervosa (BN) is thought to be characterised by reduced inhibitory control, but anorexia nervosa (AN) is thought to be characterised by increased inhibitory control (Lock, Garrett, Beenhakker, & Reiss, 2011). Therefore, an inhibitory training program that utilises negative super-learning procedures could be argued to be effective for BN, whilst a program utilising a positive super-learning procedure could be effective for AN.

Of particular relevance here is the possibility that this may provide a means for training inhibition in psychopathic individuals. As has been discussed, reduced processing capacity appears to interfere with psychopathic individuals' learning, something I have argued relates particularly to a cue's associability. I have suggested that in more severe psychopathy cases, the ability to form the initial inhibitory association may also be impaired due to its dependence on the integration of simultaneous information. In contrast, negative super-learning requires serial information processing, with an initial excitatory association being established, then using this subsequently (instead of simultaneously) to train an inhibitor. Therefore, the processing demands at any one time are lower than if an associative inhibitor is trained by simultaneously comparing two stimuli (e.g., AX-, X+). This means that negative super-learning may offer a low processing demand alternative for training inhibition in psychopathy. These suggestions are made on the basis of previous theoretical and intervention work in this area and the present results, though of course it remains for future research to determine

whether the principles established here hold true when alternative super-learning comparison cues are used, and when these findings are applied in an intervention setting.

6.8. Conclusions

In a series of experiments this thesis has explored learning and behaviour in psychopathy. Psychopathic traits have been shown to predict altered learning about inhibitory cues. A series of experimental manipulations suggest that this reflects an impaired ability to adjust a cue's associability, a consequence of a reduced information processing capacity. The inhibitory specificity of this difference is attributed to the increased processing demands linked with acquiring an inhibitory association, as compared to an excitatory association. The Psychopathy Attention Theory has been developed to account for these results. Drawing on the principles of the Impaired Integration Theory (Hamilton et al., 2015) and the Theory of Attention (Mackintosh, 1975), the Psychopathy Attention Theory reconciles observations of psychopathy differences with accounts of associative learning and represents these through a modification to Mackintosh's equation of learning. Specifically, the influence of associability on learning is argued to be inversely related to the extent of individuals' psychopathic traits. Suggestions of future experiments that could test this Psychopathy Attention Theory have been made. In addition, the novel go/nogo task developed here is argued to be of use in characterising the broad category of 'inhibitory control disorders' by establishing which aspect of inhibitory control is affected. This series of studies has also offered preliminary evidence of so-called 'negative super-learning', which has implications for inhibitory training programs, and may provide a means by which to train inhibitory control in psychopathic individuals under conditions of low information processing demands.

REFERENCES

Allan, L. G. (1980). A note on measurement of contingency between two binary variables in judgment tasks. *Bulletin of the Psychonomic Society*, *15*(3), 147–149. doi:10.3758/BF03334492

Alloy, L. B., & Abramson, L. Y. (1979). Judgment of contingency in depressed and nondepressed students: sadder but wiser? *Journal of Experimental Psychology: General*, *108*(4), 441–485. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/528910

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washing, DC. doi:10.1176/appi.books.9780890425596.893619

Baddeley, A. D., Chincotta, D., & Adlam, A. (2001). Working Memory and the Control of Action: Evidence from Task Switching. *Journal of Experimental Psychology: General*, *130*(4), 641–657.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *The Psychology of Learning and Motivation: Advances in Research and Theory* (pp. 47–89). New York: Academic Press.

Baddeley, A. D., Lewis, V., & Vallar, G. (1984). Exploring the articulatory loop. *The Quarterly Journal of Experimental Psychology Section A*, *36*(2), 233–252. doi:10.1080/14640748408402157

Baetu, I., & Baker, A. G. (2010). Extinction and blocking of conditioned inhibition in human causal learning. *Learning & Behavior*, *38*(4), 394–407. doi:10.3758/LB.38.4.394

Barkley, R. A. (1997). Behavioral inhibition, sustained attention, and executive

functions: constructing a unifying theory of ADHD. *Psychological Bulletin*, *121*(1), 65–94. doi:10.1037/0033-2909.121.1.65

Baskin-Sommers, A. R., Curtin, J. J., & Newman, J. P. (2011). Specifying the attentional selection that moderates the fearlessness of psychopathic offenders. *Psychological Science*, *22*(2), 226–234. doi:10.1177/0956797610396227

Baskin-Sommers, A. R., Curtin, J. J., & Newman, J. P. (2013). Emotion-modulated startle in psychopathy: Clarifying familiar effects. *Journal of Abnormal Psychology*, *122*(2), 458–468. doi:10.1037/a0030958.Emotion-modulated

Beesley, T., & Le Pelley, M. E. (2011). The influence of blocking on overt attention and associability in human learning. *Journal of Experimental Psychology: Animal Behavior Processes*, *37*(1), 114–120. doi:10.1037/a0019526

Benning, S. D., Patrick, C. J., Blonigen, D. M., Hicks, B. M., & Iacono, W. G. (2005). Estimating facets of psychopathy from normal personality traits: a step toward community epidemiological investigations. *Assessment*, *12*(1), 3–18. doi:10.1177/1073191104271223

Benning, S. D., Patrick, C. J., Hicks, B. M., Blonigen, D. M., & Krueger, R. F. (2003). Factor structure of the psychopathic personality inventory: validity and implications for clinical assessment. *Psychological Assessment*, *15*(3), 340–350. doi:10.1037/1040-3590.15.3.340

Birbaumer, N., & Veit, R. (2005). Deficient fear conditioning in psychopathy: a functional magnetic resonance imaging study. *Archives of General Psychiatry*, *62*(July), 799–805. Retrieved from http://archpsyc.jamanetwork.com/article.aspx?articleid=208744

Blair, K. S., Morton, J., Leonard, A., & Blair, R. J. R. (2006). Impaired decision-making on the basis of both reward and punishment information in individuals with

psychopathy. *Personality and Individual Differences*, *41*(1), 155–165. doi:10.1016/j.paid.2005.11.031

Blair, R. J. R. (2005). Applying a cognitive neuroscience perspective to the disorder of psychopathy. *Development and Psychopathology*, *17*, 865–891. doi:10.1017/S0954579405050418

Blair, R. J. R. (2007). The amygdala and ventromedial prefrontal cortex in morality and psychopathy. *Trends in Cognitive Sciences*, *11*(9), 387–392. doi:10.1016/j.tics.2007.07.003

Blair, R. J. R., & Mitchell, D. G. V. (2009). Psychopathy, attention and emotion. *Psychological Medicine*, *39*(4), 543–555. doi:10.1017/S0033291708003991

Blair, R. J. R., Mitchell, D. G. V., Leonard, A., Budhani, S., Peschardt, K. S., & Newman, C. (2004). Passive avoidance learning in individuals with psychopathy: Modulation by reward but not by punishment. *Personality and Individual Differences*, *37*(6), 1179–1192. doi:10.1016/j.paid.2003.12.001

Blair, R. J. R., Peschardt, K. S., Budhani, S., Mitchell, D. G. V., & Pine, D. S. (2006). The development of psychopathy. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *47*(3-4), 262–76. doi:10.1111/j.1469-7610.2006.01596.x

Blonigen, D. M., & Carlson, S. R. (2003). A twin study of self-reported psychopathic personality traits. *Personality and Individual …*, *35*, 179–197. Retrieved from http://www.sciencedirect.com/science/article/pii/S0191886902001848

Bonardi, C., Graham, S., Hall, G., & Mitchell, C. (2005). Acquired Equivalence and Distinctiveness in Human Discrimination Learning : Evidence for Associative Mediation. *Psychonomic Bulletin & Review*, *12*(1), 88–92. doi:10.1037/0096-3445.132.2.266

Bowditch, W. A., Verbruggen, F., & McLaren, I. P. L. (2015). Associatively mediated

stopping: Training stimulus-specific inhibitory control. *Learning & Behavior*, 1–13. doi:10.3758/s13420-015-0196-8

Bowley, C., Faricy, C., Hegarty, B., Johnstone, S. J., Smith, J. L., Kelly, P. J., & Rushby, J. A. (2013). The effects of inhibitory control training on alcohol consumption, implicit alcohol-related cognitions and brain electrical activity. *International Journal of Psychophysiology*, *89*(3), 342–348. doi:10.1016/j.ijpsycho.2013.04.011

Brazil, I. A., Hunt, L. T., Bulten, B. H., Kessels, R. P. C., de Bruijn, E. R. A., & Mars, R. B. (2013). Psychopathy-related traits and the use of reward and social information: a computational approach. *Frontiers in Psychology*, *4*(December). doi:10.3389/fpsyg.2013.00952

Brazil, I. A., Maes, J. H. R., Scheper, I., Bulten, B. H., Kessels, R. P. C., Verkes, R. J., & de Bruijn, E. R. . A. (2013). Reversal deficits in individuals with psychopathy in explicit but not implicit learning conditions. *Journal of Psychiatry & Neuroscience*, *38*(4), 13–20. doi:10.1503/jpn.120152

Brazil, I. A., Mars, R. B., Bulten, B. H., Buitelaar, J. K., Verkes, R. J., & De Bruijn, E. R. A. (2011). A neurophysiological dissociation between monitoring one's own and others' actions in psychopathy. *Biological Psychiatry*, *69*(7), 693–699. doi:10.1016/j.biopsych.2010.11.013

Budhani, S., Marsh, A. A., Pine, D. S., & Blair, R. J. R. (2007). Neural correlates of response reversal: Considering acquisition. *NeuroImage*, *34*, 1754–1765. doi:10.1016/j.neuroimage.2006.08.060

Budhani, S., Richell, R. A., & Blair, R. J. R. (2006). Impaired reversal but intact acquisition: probabilistic response reversal deficits in adult individuals with psychopathy. *Journal of Abnormal Psychology*, *115*(3), 552–558. doi:10.1037/0021-843X.115.3.552

Byrom, N. C., Msetfi, R. M., & Murphy, R. A. (2015). Two pathways to causal control: Use and availability of information in the environment in people with and without signs of depression. *Acta Psychologica*, *157*, 1–12. doi:10.1016/j.actpsy.2015.02.004

Byrom, N. C., & Murphy, R. A. (2013). It's not just what we encode, but how we encode it: associations between neuroticism and learning. *Journal of Personality*, *81*(4), 345–354. doi:10.1111/jopy.12022

Cale, E. M., & Lilienfeld, S. O. (2002). Sex differences in psychopathy and antisocial personality disorder: A review and integration. *Clinical Psychology Review*, *22*, 1179–1207. doi:10.1016/S0272-7358(01)00125-8

Chatlosh, D. L., Neunaber, D. J., & Wasserman, E. A. (1985). Response-outcome contingency: Behavioral and judgmental effects of appetitive and aversive outcomes with college students. *Learning and Motivation*, *16*(1), 1–34. doi:10.1016/0023-9690(85)90002-5

Clark, L., Cools, R., & Robbins, T. W. (2004). The neuropsychology of ventral prefrontal cortex: Decision-making and reversal learning. *Brain and Cognition*, *55*(1), 41–53. doi:10.1016/S0278-2626(03)00284-7

Cleckley, H. (1982). *The Mask of Sanity. Revised Edition*. Saint Louis, MO: Mosby Medical Library.

Cooke, D. J., & Michie, C. (2001). Refining the construct of psychopathy: towards a hierarchical model. *Psychological Assessment*, *13*(2), 171–188. doi:10.1037/1040-3590.13.2.171

Cools, R., Clark, L., Owen, A. M., & Robbins, T. W. (2002). Defining the neural mechanisms of probabilistic reversal learning using event-related functional magnetic resonance imaging. *The Journal of Neuroscience*, *22*(11), 4563–4567.

doi:20026435

Corr, P. J. (2002). J. A. Gray's reinforcement sensitivity theory: Tests of the joint subsystems hypothesis of anxiety and impulsivity. *Personality and Individual Differences*, *33*(4), 511–532. doi:10.1016/S0191-8869(01)00170-2

Darredeau, C., Baetu, I., Baker, A. G., & Murphy, R. A. (2009). Competition between multiple causes of a single outcome in causal reasoning. *Journal of Experimental Psychology: Animal Behavior Processes*, *35*(1), 1–14. doi:10.1037/a0012699

De Houwer, J., & Beckers, T. (2003). Secondary task difficulty modulates forward blocking in human contingency learning. *The Quarterly Journal of Experimental Psychology B: Comparative and Physiological Psychology*, *56*(4), 345–357. doi:10.1080/02724990244000296

De Wit, S., & Dickinson, A. (2009). Associative theories of goal-directed behaviour: A case for animal-human translational models. *Psychological Research*, *73*(4), 463–476. doi:10.1007/s00426-009-0230-6

Deubel, H., & Schneider, W. (1996). Saccade target selection and object recognition: Evidence for a common attentional mechanism. *Vision Research*, *36*(12), 1827–1837. Retrieved from http://linkinghub.elsevier.com/retrieve/pii/0042698995002944

Dickinson, A. (1985). Actions and Habits: The Development of Behavioural Autonomy. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *308*, 67–78. doi:10.1098/rstb.1985.0010

Dickinson, A., Shanks, D. R., & Evenden, J. (1984). Judgment of Act-Outcome Contingency - the Role of Selective Attribution. *The Quarterly Journal of Experimental Psychology Section A*, *36*(1), 29–50. doi:10.1080/14640748408401502

Driver, J. (2001). A selective review of selective attention research from the past century.

British Journal of Psychology, *92*, 53–78. doi:10.1348/000712601162103

Dube, W., Lombard, K., Farren, K., Flusser, D., Balsamo, L., & Fowler, T. (1999). Eye tracking assessment of stimulus overselectivity in individuals with mental retardation. *Experimental Analysis of Human Behavior Bulletin*, *17*, 8–14. Retrieved from http://eahb.org/Dubeetal/Dubeetal.pdf

Dube, W., & McIlvane, W. (1999). Reduction of stimulus overselectivity with nonverbal differential observing responses. *Journal of Applied Behavior Analysis*, *32*(1), 25–33. doi:10.1901/jaba.1999.32-25

Dunbarb, K., & Sussman, D. (1995). Toward a Cognitive Account of Frontal Lobe Function - Simulating Frontal Lobe Deficits in Normal Subjects. *Annals of the New York Academy of Sciences*, *769*, 289–304.

Eysenck, M. W. (1988). Anxiety and attention. *Anxiety Research*, *1*(1), 9–15. doi:10.1080/10615808808248216

Falkenstein, M., Hoormann, J., & Hohnsbein, J. (1999). ERP components in Go/Nogo tasks and their relation to inhibition. *Acta Psychologica*, *101*, 267–291. doi:10.1016/S0001-6918(99)00008-6

Flor, H., Birbaumer, N., Hermann, C., Ziegler, S., & Patrick, C. J. (2002). Aversive Pavlovian conditioning in psychopaths: peripheral and central correlates. *Psychophysiology*, *39*(4), 505–518. doi:10.1017.S0048577202394046

Fox, E. (1993). Allocation of visual attention and anxiety. *Cognition & Emotion*, *7*(2), 207–215. doi:10.1080/02699939308409185

Gamer, M., & Büchel, C. (2009). Amygdala Activation Predicts Gaze toward Fearful Eyes. *The Journal of Neuroscience*, *29*(28), 9123–9126. doi:10.1523/JNEUROSCI.1883-09.2009

Gao, Y., & Raine, A. (2010). Successful and unsuccessful psychopaths: A

neurobiological model. *Behavioral Sciences and the Law*, *28*(2), 194–210. doi:10.1002/bsl.924

Gao, Y., Raine, A., Venables, P. H., Dawson, M. E., & Mednick, S. A. (2010). Association of poor childhood fear conditioning and adult crime. *The American Journal of Psychiatry*, *167*(1), 56–60. doi:10.1176/appi.ajp.2009.09040499

Gillan, C. M., Morein-Zamir, S., Durieux, A. M. S., Fineberg, N. A., Sahakian, B. J., & Robbins, T. W. (2014). Obsessive-compulsive disorder patients have a reduced sense of control on the illusion of control task. *Frontiers in Psychology*, *5*(March), 204. doi:10.3389/fpsyg.2014.00204

Gorenstein, E. E., & Newman, J. P. (1980). Disinhibitory psychopathology: a new perspective and a model for research. *Psychological Review*, *87*(3), 301–315. doi:doi:10.1037/0033-295x.87.3.301

Gregory, S., Blair, R. J. R., Ffytche, D., Simmons, A., Kumari, V., Hodgins, S., & Blackwood, N. (2015). Punishment and psychopathy: a case-control functional MRI investigation of reinforcement learning in violent antisocial personality disordered men. *The Lancet Psychiatry*, *2*(2), 153–160. doi:10.1016/S2215-0366(14)00071-6

Gullo, M. J., Jackson, C. J., & Dawe, S. (2010). Impulsivity and reversal learning in hazardous alcohol use. *Personality and Individual Differences*, *48*(2), 123–127. doi:10.1016/j.paid.2009.09.006

Hall, G., Mackintosh, N. J., Goodall, G., & Martello, M. D. (1977). Loss of control by a less valid or by a less salient stimulus compounded with a better predictor of reinforcement. *Learning and Motivation*, *8*(2), 145–158. doi:10.1016/0023-9690(77)90001-7

Hamilton, R. K. B., Hiatt Racer, K., & Newman, J. P. (2015). Impaired integration in psychopathy: A unified theory of psychopathic dysfunction. *Psychological Review*,

122(4), 770–791. doi:10.1037/a0039703

Hare, R. D. (1991). *The Hare Psychopathy Checklist-Revised (PCL-R)*. Toronto, Ontario: Multi-Health Systems, Incorporated.

Hare, R. D., Hart, S. D., & Harpur, T. J. (1991). Psychopathy and the DSM-IV criteria for antisocial personality disorder. *Journal of Abnormal Psychology*, *100*(3), 391–398. doi:10.1037/0021-843X.100.3.391

Hare, R. D., & Neumann, C. S. (2008). Psychopathy as a Clinical and Empirical Construct. *Annual Review of Clinical Psychology*, *4*(1), 217–246. doi:10.1146/annurev.clinpsy.3.022806.091452

Harpur, T. J., Hare, R. D., & Hakstian, A. R. (1989). Two-factor conceptualization of psychopathy: Construct validity and assessment implications. *Psychological Assessment*, *1*(1), 6–17. doi:10.1037//1040-3590.1.1.6

Hart, S. D., & Dempster, R. J. (1997). Impulsivity and Psychopathy. In C. D. Webster & M. A. Jackson (Eds.), *Impulsivity: Theory, assessment and treatment* (pp. 212–232). New York: Guildford Press.

He, Z., Cassaday, H. J., Howard, R. C., Khalifa, N., & Bonardi, C. (2011). Impaired Pavlovian conditioned inhibition in offenders with personality disorders. *Quarterly Journal of Experimental Psychology*, *64*(12), 2334–2351. doi:10.1080/17470218.2011.616933

Hemphill, J. F., Hare, R. D., & Wong, S. (1998). Psychopathy and recidivism: A review. *Legal and Criminological Psychology*, *3*(1), 139–170.

Hiatt, K. D., Schmitt, W. A., & Newman, J. P. (2004). Stroop tasks reveal abnormal selective attention among psychopathic offenders. *Neuropsychology*, *18*(1), 50–59. doi:10.1037/0894-4105.18.1.50

Hoppenbrouwers, S. S., Van der Stigchel, S., Slotboom, J., Dalmaijer, E. S., &

Theeuwes, J. (2015). Disentangling attentional deficits in psychopathy using visual search: Failures in the use of contextual information. *Personality and Individual Differences*, *86*, 132–138. doi:10.1016/j.paid.2015.06.009

Houben, K. (2011). Overcoming the urge to splurge: Influencing eating behavior by manipulating inhibitory control. *Journal of Behavior Therapy and Experimental Psychiatry*, *42*(3), 384–388. doi:10.1016/j.jbtep.2011.02.008

Iaboni, F., Douglas, V. I., & Baker, A. G. (1995). Effects of Reward and Response Costs on Inhibition in ADHD Children. *Journal of Abnormal Psychology*, *104*(1), 232–240. Retrieved from http://psycnet.apa.org/psycinfo/1995-17950-001

Ishikawa, S. S., Raine, A., Lencz, T., Bihrle, S., & Lacasse, L. (2001). Autonomic stress reactivity and executive functions in successful and unsuccessful criminal psychopaths from the community. *Journal of Abnormal Psychology*, *110*(3), 423–432. doi:10.1037/0021-843X.110.3.423

Jones, A., & Field, M. (2013). The effects of cue-specific inhibition training on alcohol consumption in heavy social drinkers. *Experimental and Clinical Psychopharmacology*, *21*(1), 8–16. doi:10.1037/a0030683

Kamarajan, C., Porjesz, B., Jones, K. A., Choi, K., Chorlian, D. B., Padmanabhapillai, A., … Begleiter, H. (2005). Alcoholism is a Disinhibitory Disorder: Neuropsychological Evidence from a Go/No-Go Task. *Biological Psychology*, *69*(3), 353–373. doi:10.1016/j.biopsycho.2004.08.004.Alcoholism

Kamin, L. J. (1969). Predictability, surprise, attention and conditioning. In B. A. Campbell & R. M. Church (Eds.), *Punishment and Aversive Behaviour* (pp. 279–296). New York: Appleton-Century-Crofts.

Kaufman, J. N., Ross, T. J., Stein, E. A., & Garavan, H. (2003). Cingulate hypoactivity in cocaine users during a GO-NOGO task as revealed by event-related functional

magnetic resonance imaging. *The Journal of Neuroscience*, *23*(21), 7839–7843. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12944513

Kiehl, K. A., Smith, A. M., Hare, R. D., & Liddle, P. F. (2000). An event-related potential investigation of response inhibition in schizophrenia and psychopathy. *Biological Psychiatry*, *48*(3), 210–221. doi:10.1016/s0006-3223(00)00834-9

Lapierre, D., Braun, C. M. J., & Hodgins, S. (1995). Ventral frontal deficits in psychopathy: Neuropsychological test findings. *Neuropsychologia*, *33*(2), 139–151. doi:10.1016/0028-3932(95)98935-6

Le Pelley, M. E. (2004). The role of associative history in models of associative learning: a selective review and a hybrid model. *The Quarterly Journal of Experimental Psychology. B, Comparative and Physiological Psychology*, *57*(3), 193–243. doi:10.1080/02724990344000141

Le Pelley, M. E., Beesley, T., & Griffiths, O. (2011). Overt attention and predictiveness in human contingency learning. *Journal of Experimental Psychology: Animal Behavior Processes*, *37*(2), 220–229. doi:10.1037/a0021384

Le Pelley, M. E., Beesley, T., & Griffiths, O. (2014). Relative salience versus relative validity: cue salience influences blocking in human associative learning. *Journal of Experimental Psychology: Animal Learning and Cognition*, *40*(1), 116–132. doi:10.1037/xan0000006

Le Pelley, M. E., & McLaren, I. P. L. (2003). Learned associability and associative change in human causal learning. *The Quarterly Journal of Experimental Psychology*, *56*(1), 68–79. doi:10.1080/02724990244000179

Le Pelley, M. E., Mitchell, C. J., & Johnson, A. M. (2013). Outcome value influences attentional biases in human associative learning: Dissociable effects of training and instruction. *Journal of Experimental Psychology: Animal Behavior Processes*,

39(1), 39–55. doi:10.1037/a0031230

Le Pelley, M. E., Oakeshott, S. M., & McLaren, I. P. L. (2005). Blocking and unblocking in human causal learning. *Journal of Experimental Psychology: Animal Behavior Processes*, *31*(1), 56–70. doi:10.1037/0097-7403.31.1.56

Lee, I. A., & Preacher, K. J. (2013). Calculation for the test of the difference between two dependent correlations with one variable in common [Computer software].

Lee, Z., & Salekin, R. T. (2010). Psychopathy in a noninstitutional sample: differences in primary and secondary subtypes. *Personality Disorders*, *1*(3), 153–169. doi:10.1037/a0019269

Lilienfeld, S. O., Patrick, C. J., Benning, S. D., Berg, J., Sellbom, M., & Edens, J. F. (2012). The role of fearless dominance in psychopathy: Confusions, controversies, and clarifications. *Personality Disorders: Theory, Research, and Treatment*, *3*(3), 327–340. doi:10.1037/a0026987

Lilienfeld, S. O. S., & Hess, T. T. H. T. (2001). Psychopathic personality traits and somatization: Sex differences and the mediating role of negative emotionality. *Journal of Psychopathology and Behavioral …*, *23*(1), 11–24. doi:10.1023/A:1011035306061

Lilienfeld, S. O., & Widows, M. R. (2005). *Psychopathic Personality Inventory - Revised: Professional Manual*. Lutz, FL: Psychological Assessment Resources.

Livesley, W. J. (2007). A framework for integrating dimensional and categorical classifications of personality disorders. *Journal of Personality Disorders*, *21*(2), 199–244.

Lock, J., Garrett, A., Beenhakker, J., & Reiss, A. (2011). Abberant Brain Activation During a Response Inhibition Task in Adolexcent Eating Disorder Subtypes. *American Journal of Psychiatry*, *168*(1), 55–64.

doi:10.1176/appi.ajp.2010.10010056.Aberrant

López, R., Poy, R., Patrick, C. J., & Moltó, J. (2013). Deficient fear conditioning and self-reported psychopathy: The role of fearless dominance. *Psychophysiology*, *50*(2), 210–218. doi:10.1111/j.1469-8986.2012.01493.x

Lovibond, P. F., Siddle, D. A. T., & Bond, N. (1988). Insensitivity to stimulus validity in human pavlovian conditioning. *The Quarterly Journal of Experimental Psychology Section B: Comparative and Physiological Psychology*, *40B*(4), 377–410. doi:10.1080/14640748808402331

Lubow, R. E. (1973). Latent Inhibition. *Psychological Bulletin*, *79*(6), 398–407. Retrieved from http://psycnet.apa.org/journals/bul/79/6/398/

Lykken, D. T. (1995). The antisocial personalities. Hilsdale, NJ: Erlbaum.

MacCoon, D. G., Wallace, J. F., & Newman, J. P. (2004). Self-regulation: Context-Appropriate Balanced Attention. In *Handbook of Self-Regulation Research* (pp. 422–444). New York: Guildford Press.

Mackintosh, N. J. (1975). A theory of attention: Variations in the associability of stimuli with reinforcement. *Psychological Review*, *82*(4), 276–298. doi:10.1037/h0076778

Mackintosh, N. J., & Holgate, V. (1968). Effects of inconsistent reinforcement on reversal and nonreversal shifts. *Journal of Experimental Psychology*, *76*(1), 154–159. doi:10.1037/h0025310

Mahmut, M. K., Homewood, J., & Stevenson, R. J. (2008). The characteristics of non-criminals with high psychopathy traits: Are they similar to criminal psychopaths? *Journal of Research in Personality*, *42*, 679–692. doi:10.1016/j.jrp.2007.09.002

Matute, H. (1996). Illusion of Control: Detecting Response-Outcome Independence in Analytic but Not in Naturalistic Conditions. *Psychological Science*, *7*(5), 289–293. doi:10.1111/j.1467-9280.1996.tb00376.x

McLaren, I. P. L., & Verbruggen, F. (2015). Association and Inhibition. In R. A. Murphy & R. C. Honey (Eds.), *The Wiley Blackwell Handbook on The Cognitive Neuroscience of Learning*. Chichester: Wiley.

Melchers, K. G., Wolff, S., & Lachnit, H. (2006). Extinction of conditioned inhibition through nonreinforced presentation of the inhibitor. *Psychonomic Bulletin & Review*, *13*(4), 662–7. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/17201367

Menon, V., Adleman, N. E., White, C. D., Glover, G. H., & Reiss, A. L. (2001). Error-Related Brain Activation during a Go/NoGo Response Inhibition Task. *Human Brain Mapping*, *12*, 131–143. doi:10.1002/1097-0193(200103)12:3<131::AID-HBM1010>3.0.CO;2-C

Miller, J. D., & Lynam, D. R. (2012). An examination of the Psychopathic Personality Inventory's nomological network: A meta-analytic review. *Personality Disorders: Theory, Research, and Treatment*, *3*(3), 305–326. doi:10.1037/a0024567

Mitchell, D. G. V., Colledge, E., Leonard, A., & Blair, R. J. R. (2002). Risky decisions and response reversal: Is there evidence of orbitofrontal cortex dysfunction in psychopathic individuals? *Neuropsychologia*, *40*(12), 2013–2022. doi:10.1016/S0028-3932(02)00056-8

Moeller, F. G., Barratt, E. S., Dougherty, D. M., Schmitz, J. M., & Swann, A. C. (2001). Psychiatric aspects of impulsivity. *American Journal of Psychiatry*, *158*(11), 1783–1793. doi:10.1176/appi.ajp.158.11.1783

Moul, C., Killcross, S., & Dadds, M. R. (2012). A model of differential amygdala activation in psychopathy. *Psychological Review*, *119*(4), 789–806. doi:10.1037/a0029342

Msetfi, R. M., Murphy, R. A., & Simpson, J. (2007). Depressive realism and the effect of

intertrial interval on judgements of zero, positive, and negative contingencies. *Quarterly Journal of Experimental Psychology*, *60*(3), 461–481. doi:10.1080/17470210601002595

Msetfi, R. M., Murphy, R. A., Simpson, J., & Kornbrot, D. E. (2005). Depressive realism and outcome density bias in contingency judgments: the effect of the context and intertrial interval. *Journal of Experimental Psychology: General*, *134*(1), 10–22. doi:10.1037/0096-3445.134.1.10

Msetfi, R. M., Wade, C., & Murphy, R. A. (2013). Context and time in causal learning: contingency and mood dependent effects. *PloS One*, *8*(5), e64063. doi:10.1371/journal.pone.0064063

Munro, G. E. S., Dywan, J., Harris, G. T., McKee, S., Unsal, A., & Segalowitz, S. J. (2007a). ERN varies with degree of psychopathy in an emotion discrimination task. *Biological Psychology*, *76*, 31–42. doi:10.1016/j.biopsycho.2007.05.004

Munro, G. E. S., Dywan, J., Harris, G. T., McKee, S., Unsal, A., & Segalowitz, S. J. (2007b). Response inhibition in psychopathy: the frontal N2 and P3. *Neuroscience Letters*, *418*(2), 149–153. doi:10.1016/j.neulet.2007.03.017

Murphy, R. a, Schmeer, S., Vallée-Tourangeau, F., Mondragón, E., & Hilton, D. (2011). Making the illusory correlation effect appear and then disappear: the effects of increased learning. *Quarterly Journal of Experimental Psychology (2006)*, *64*(1), 24–40. doi:10.1080/17470218.2010.493615

Musca, S. C., Vadillo, M. A., Blanco, F., & Matute, H. (2010). The role of cue information in the outcome-density effect: evidence from neural network simulations and a causal learning experiment. *Connection Science*, *22*(2), 177–192. doi:10.1080/09540091003623797

Neumann, D. L., & Waters, A. M. (2006). The use of an unpleasant sound as an

unconditional stimulus in human aversive Pavlovian conditioning procedure. *Biological Psychology*, *73*, 175–185. doi:10.1016/j.biopsycho.2008.08.005

Neumann, D. L., Waters, A. M., & Westbury, H. R. (2008). The use of an unpleasant sound as the unconditional stimulus in aversive Pavlovian conditioning experiments that involve children and adolescent participants. *Behavior Research Methods*, *40*(2), 622–625. doi:10.3758/BRM.40.2.622

Newman, J. P. (1998). Psychopathic Behaviour: An information processing perspective. In *Psychopathy: Theory, Research and Implications for Society.* (pp. 81–104). Netherlands: Springer.

Newman, J. P., & Baskin-Sommers, A. R. (2011). Early Selective Attention Abnormalities in Psychopathy. In M. I. Posner (Ed.), *Cognitive Neuroscience of Attention* (2nd ed., pp. 421–440). Guildford Publications.

Newman, J. P., Curtin, J. J., Bertsch, J. D., & Baskin-Sommers, A. R. (2010). Attention moderates the fearlessness of psychopathic offenders. *Biological Psychiatry*, *67*(1), 66–70. doi:10.1016/j.biopsych.2009.07.035

Newman, J. P., & Kosson, D. S. (1986). Passive Avoidance Learning in Psychopathic and Nonpsychopathic Offenders. *Journal of Abnormal Psychology*, *95*(3), 252–256. doi:doi:10.1037/0021-843x.95.3.252

Newman, J. P., Patterson, C. M., Howland, E. W., & Nichols, S. L. (1990). Passive Avoidance in Psychopaths: The Effects of Reward. *Personality and Individual Differences*, *11*(11), 1101–1114. doi:10.1016/0191-8869(90)90021-i

Newman, J. P., Patterson, C. M., & Kosson, D. S. (1987). Response perseveration in psychopaths. *Journal of Abnormal Psychology*, *96*(2), 145–148. doi:10.1037/0021-843x.96.2.145

Newman, J. P., & Schmitt, W. A. (1998). Passive avoidance in psychopathic offenders: a

replication and extension. *Journal of Abnormal Psychology*, *107*(3), 527–532. doi:10.1037/0021-843X.107.3.527

Newman, J. P., Widom, C. S., & Nathan, S. (1985). Passive avoidance in syndromes of disinhibition: psychopathy and extraversion. *Journal of Personality and Social Psychology*, *48*(5), 1316–1327. doi:10.1037/0022-3514.48.5.1316

Nigg, J. T. (2000). On inhibition/disinhibition in developmental psychopathology: Views from cognitive and personality psychology and a working inhibition taxonomy. *Psychological Bulletin*, *126*(2), 220–246. doi:10.1037//0033-2909.126.2.220

Nigg, J. T., Wong, M. M., Martel, M. M., Jester, J. M., Puttler, L. I., Glass, J. M., … Zucker, R. A. (2006). Poor Response Inhibition as a Predictor of Problem Drinking and Illicit Drug Use in Adolescents at Risk for Alcoholism and Other Substance Use Disorders. *Journal of the American Academy of Child & Adolescent Psychiatry*, *45*(4), 468–475. doi:10.1097/01.chi.0000199028.76452.a9

Oades, R. D., Lasky-Su, J., Christiansen, H., Faraone, S. V., Sonuga-Barke, E. J. S., Banaschewski, T., … Asherson, P. (2008). The influence of serotonin- and other genes on impulsive behavioral aggression and cognitive impulsivity in children with attention-deficit/hyperactivity disorder (ADHD): Findings from a family-based association test (FBAT) analysis. *Behavioral and Brain Functions*, *4*(1), 48–60. doi:10.1186/1744-9081-4-48

Patrick, C. J. (2006). Back to the future: Cleckley as a guide to the next generation of psychopathy research. In *Handbook of Psychopathy* (pp. 605–617). New York: Guildford Press.

Patrick, C. J., Fowles, D. C., & Krueger, R. F. (2009). Triarchic conceptualization of psychopathy: developmental origins of disinhibition, boldness, and meanness. *Development and Psychopathology*, *21*(3), 913–938.

doi:10.1017/S0954579409000492

Patterson, C. M., & Newman, J. P. (1993). Reflectivity and learning from aversive events: toward a psychological mechanism for the syndromes of disinhibition. *Psychological Review*, *100*, 716–736. doi:10.1037/0033-295X.100.4.716

Pearce, J. M., & Hall, G. (1980). A model for Pavlovian learning: variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychological Review*, *87*(6), 532–552. doi:10.1037/0033-295X.87.6.532

Perry, J. L., & Carroll, M. E. (2008). The role of impulsive behavior in drug abuse. *Psychopharmacology*, *200*(1), 1–26. doi:10.1007/s00213-008-1173-0

Poythress, N. G., & Hall, J. R. (2011). Psychopathy and impulsivity reconsidered. *Aggression and Violent Behavior*, *16*(2), 120–134. doi:10.1016/j.avb.2011.02.003

Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods*, *40*(3), 879–891. doi:10.3758/BRM.40.3.879

Rescorla, R. A. (1969). Pavlovian Conditioned Inhibition. *Psychological Bulletin*, *72*(2), 77–94. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/19586159

Rescorla, R. A. (1971a). Summation and retardation tests of latent inhibition. *Journal of Comparative and Physiological Psychology*, *75*(1), 77–81. doi:http://dx.doi.org/10.1037/h0030694

Rescorla, R. A. (1971b). Variation in the Effectiveness of Reinforcement Prior and Nonreinforcement Inhibitory Following. *Learning and Motivation*, *2*, 113–123.

Rescorla, R. A. (1990). The role of information about the response-outcome relation in instrumental discrimination learning. *Journal of Experimental Psychology: Animal Behavior Processes*, *16*(3), 262–270. doi:10.1037//0097-7403.16.3.262

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations

in the effectiveness of reinforcement and nonreinforcement. In B. A.H. & Prokasy W.F. (Eds.), *Classical Conditioning II: Current Research and Theory* (pp. 64–99). New York: Appleton Century Crofts.

Rodriguez, G., & Hall, G. (2008). Potentiation of latent inhibition. *Journal of Experimental Psychology: Animal Behavior Processes*, *34*(3), 352–360. doi:10.1037/0097-7403.34.3.352

Sadeh, N., & Verona, E. (2008). Psychopathic personality traits associated with abnormal selective attention and impaired cognitive control. *Neuropsychology*, *22*(5), 669–680. doi:10.1037/a0012692

Salekin, R. T., Rogers, R., & Sewell, K. W. (1996). A review and meta-analysis of the Psychopathy Checklist and Psychopathy Checklist-Revised: Predictive validity of dangerousness. *Clinical Psychology: Science and Practice*, *3*(3), 203–215. Retrieved from http://onlinelibrary.wiley.com/doi/10.1111/j.1468-2850.1996.tb00071.x/full

Shanks, D. R. (2007). Associationism and cognition: human contingency learning at 25. *Quarterly Journal of Experimental Psychology*, *60*(3), 291–309. doi:10.1080/17470210601000581

Swann, A. C., Anderson, J. C., Dougherty, D. M., & Moeller, F. G. (2001). Measurement of inter-episode impulsivity in bipolar disorder. *Psychiatry Research*, *101*(2), 195–197. doi:10.1016/S0165-1781(00)00249-3

Swann, A. C., Pazzaglia, P., Nicholls, A., Dougherty, D. M., & Moeller, F. G. (2003). Impulsivity and phase of illness in bipolar disorder. *Journal of Affective Disorders*, *73*, 105–111. doi:10.1016/S0165-0327(02)00328-2

Thomas, D. R. (1970). Stimulus selection, attention and related matters. In J. H. Reynierse (Ed.), *Current Issues in Animal Learning* (pp. 311–356). Lincoln:

University of Nebraska Press.

Torrubia, R., Ávila, C., Moltó, J., & Caseras, X. (2001). The Sensitivity to Punishment and Sensitivity to Reward Questionnaire (SPSRQ) as a measure of Gray's anxiety and impulsivity dimensions. *Personality and Individual Differences*, *31*(6), 837–862. doi:10.1016/S0191-8869(00)00183-5

Tullett, A. M., & Inzlicht, M. (2010). The voice of self-control: Blocking the inner voice increases impulsive responding. *Acta Psychologica*, *135*(2), 252–256. doi:10.1016/j.actpsy.2010.07.008

Vallée-Tourangeau, F., Murphy, R. A., Drew, S., & Baker, A. G. (1998). Judging the Importance of Constant and Variable Candidate Causes: A Test of the Power PC Theory. *The Quarterly Journal of Experimental Psychology*, *51A*(1), 65–84.

Veling, H., Aarts, H., & Stroebe, W. (2013). Using stop signals to reduce impulsive choices for palatable unhealthy foods. *British Journal of Health Psychology*, *18*(2), 354–368. doi:10.1111/j.2044-8287.2012.02092.x

Verbruggen, F., Chambers, C. D., & Logan, G. D. (2013). Fictitious inhibitory differences: how skewness and slowing distort the estimation of stopping latencies. *Psychological Science*, *24*(3), 352–362. doi:10.1177/0956797612457390

Verbruggen, F., Logan, G. D., & Stevens, M. A. (2008). STOP-IT: Windows executable software for the stop-signal paradigm. *Behavior Research Methods*, *40*(2), 479–483. doi:10.3758/BRM.40.2.479

Verona, E., Patrick, C. J., & Joiner, T. E. (2001). Psychopathy, antisocial personality, and suicide risk. *Journal of Abnormal Psychology*, *110*(3), 462–470. doi:10.1037/0021-843X.110.3.462

Verona, E., Sprague, J., & Sadeh, N. (2012). Inhibitory control and negative emotional processing in psychopathy and antisocial personality disorder. *Journal of Abnormal*

Psychology, *121*(2), 498–510. doi:10.1037/a0025308

von Borries, A. K. L., Brazil, I. A., Bulten, B. H., Buitelaar, J. K., Verkes, R. J., & de Bruijn, E. R. A. (2010). Neural correlates of error-related learning deficits in individuals with psychopathy. *Psychological Medicine*, *40*(9), 1559–1568. doi:10.1017/S0033291709992017

Wagner, A. R. (1969). Stimulus validity and stimulus selection in associative learning. In N. J. Mackintosh & W. K. Honig (Eds.), *Fundamental Issues in associative learning* (pp. 90–122). Halifax: Dalhousie University Press.

Wagner, A. R., Logan, F. A., & Haberlandt, K. (1968). Stimulus Selection in Animal Discrimination and Learning. *Journal of Experimental Psychology*, *76*(2), 171–180. Retrieved from http://psycnet.apa.org/psycinfo/1968-06983-001

Wasserman, E. A., Elek, S. M., Chatlosh, D. L., & Baker, A. G. (1993). Rating causal relations: Role of probability in judgments of response-outcome contingency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(1), 174–188. doi:10.1037/0278-7393.19.1.174

Waters, A. M., Nitz, A. B., Craske, M. G., & Johnson, C. (2007). The effects of anxiety upon attention allocation to affective stimuli. *Behaviour Research and Therapy*, *45*(4), 763–774. doi:10.1016/j.brat.2006.07.002

Williams, B. A., & McDevitt, M. A. (2002). Inhibition and Superconditioning. *Psychological Science*, *13*(5), 454–459. doi:10.1111/1467-9280.00480

Williams, D. A., Mehta, R., & Dumont, J.-L. (2004). Conditions favoring superconditioning of irrelevant conditioned stimuli. *Journal of Experimental Psychology: Animal Behavior Processes*, *30*(2), 148–159. doi:10.1037/0097-7403.30.2.148

Winstanley, C. A., Eagle, D. M., & Robbins, T. W. (2006). Behavioral models of

impulsivity in relation to ADHD: translation between clinical and preclinical studies. *Clinical Psychology Review*, *26*(4), 379–395. doi:10.1016/j.cpr.2006.01.001

Wolf, R. C., Carpenter, R. W., Warren, C. M., Zeier, J. D., Baskin-Sommers, A. R., & Newman, J. P. (2012). Reduced susceptibility to the attentional blink in psychopathic offenders: Implications for the attention bottleneck hypothesis. *Neuropsychology*, *26*(1), 102–109. doi:10.1037/a0026000